Foundations of Abstract analysis

Nishant Panda

2451 S. Timberline Rd, Apt 4—207, Fort Collins, Co 80525
E-mail address: nishant.panda@gmail.com

2010 *Mathematics Subject Classification.* Primary

Key words and phrases. measure theory, probability, integration

Contents

Part 1

Metric Spaces

Chapter 1

Metric spaces

In this chapter we will start with the description of metric spaces and some topological concepts involved in such spaces. Metrics are abstraction of the concept of distance that is familiar to us.

1.1. Basic definitions

Definition 1.1.1. *A metric space is a pair (X, d) where X is a non-empty set and d is a function*

$$d \ : X \times X \to [0, \infty),$$

called a metric, that satisfies the following properties for all x, y, z in X:

- *(positive reflexive) $d(x, y) = 0$ if and only if $x = y$.*
- *(symmetric) $d(x, y) = d(y, x)$.*
- *(triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$.*

These properties are minimal in order to capture the idea of distance between two points.

Example 1.1.1. *We will show some basic examples of a metric space.*

(1) *Let $X = \mathbb{R}$ and define $d(x, y) = |x - y|$. This is the standard metric in $\mathbb{R}$.*
(2) *Let $X = \mathbb{R}^n$ and for $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$ define*

$$d(\boldsymbol{x}, \boldsymbol{y}) = (\sum_{j=1}^{n} (x_j - y_j)^2)^{\frac{1}{2}}.$$

*The proof that d is a metric relies on the famous **Cauchy-Schwarz** inequality. This metric is called the **Euclidean** metric and in $\mathbb{R}^2$ gives us the standard Pythagorean distance.*

(3) *Let $X = \mathbb{R}^n$ and for $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$ define*

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{n} |x_j - y_j|.$$

To verify that (X, d) is indeed a metric space we need to check if d is a metric. The hardest one to check is triangle inequality. But this follwows from the triangle inequality of absolute values in $\mathbb{R}$ i.e. for any z_j in $\mathbb{R}$, $|x_j - z_j| \leq |x_j - y_j| + |y_j - z_j|$.

(4) *Let $X = \mathbb{R}^n$ and for $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$ define,*

$$d(\boldsymbol{x}, \boldsymbol{y}) = \max \{|x_j - y_j| \ : \ 1 \leq j \leq n\}.$$

(5) *Any non-empty set can be made into a metric space. Let X be a non-empty set and for any $x, y \in X$ define*

$$d(x, y) = \begin{cases} 0 & \textit{if } x = y, \\ 0 & \textit{if } x \neq y. \end{cases}$$

*This is easily verified to be a metric space and d is called the **discrete** metric.*

(6) *Suppose (X, d) is given. If Y is a nonempty subset of X, then (Y, d) is a metric space and is referred to as a subspace. As a specific example consider the set $X = \mathbb{R}$ and the metric space $(\mathbb{R}, d)$ where d is any metric. Then $Y = [0, 1]$ defines a metric space (Y, d).*

Now that we have a notion of distance we can define a notion of **closeness**. This is done through the concept of *balls* in a metric space.

Definition 1.1.2 (Open Ball)**.** *Let (X, d) be a metric space and let x be an element of X and r be a positive real number. The set*

$$\mathcal{B}_r(x) = \{y \in X \ : \ d(y, x) < r\},$$

is called the open ball of radius r around x.

Definition 1.1.3 (Closed Ball)**.** *Let (X, d) be a metric space and let x be an element of X and r be a positive real number. The set*

$$\overline{\mathcal{B}_r(x)} = \{y \in X \ : \ d(y, x) \leq r\},$$

is called the closed ball of radius r around x.

Notice that when $s < r$, $\overline{\mathcal{B}_s(x)} \subset \mathcal{B}_r(x)$ for any metric. It is highly illustrative to figure out unit balls in a metric space. For example, let us take

$X = \mathbb{R}^2$ and consider two metric spaces (X, d_1) and (X, d_2) where d_1 is the Euclidean metric defined as in 1.1.1 (2) and d_2 be the metric defined as in 1.1.1 (3). How do the open balls of radius 1 *look* around the point $\boldsymbol{x} = \mathbf{0}$ in these two metric spaces?

The notion of closeness can be made precise with the concept of *open* sets in metric spaces.

Definition 1.1.4 (Open set)**.** *If (X, d) is a metric space, then a subset G of X is open if for each x in G there is a radius $r > 0$ such that the ball of radius r around x is contained entirely in G. This is denoted logically as,*

$$\forall\,(x \in G)\,\exists\,(r > 0 \in \mathbb{R})\,[\mathcal{B}_r\,(x) \subset\ G]\,.$$

When is a set $G \subset X$ **NOT** open? If we can find an $x \in G$ such that no matter what $r > 0$ we choose, the open ball of radius r around x is not contained in G. This is immediately evident by negating the logical statement in the definition.

Definition 1.1.5 (Closed set)**.** *If (X, d) is a metric space, then a subset F of X is closed if the complement of F in X is open i.e. $X - F$ is open. This is denoted logically as,*

$$\forall\,(x \in X)\,[x \notin F \implies \exists\,(r > 0 \in \mathbb{R})\,\mathcal{B}_r\,(x) \cap F = \emptyset]\,.$$

The above definition can be made entirely in terms of the set F by taking the contrapositive of the implication in the logical statement above i.e. a set F is closed if for any point $x \in X$ it is the case that no matter what radius r we take the ball of radius r around x always intersects F, then x must be in F. Logically,

$$\forall\,(x \in X)\,[\forall\,(r > 0 \in \mathbb{R})\,\mathcal{B}_r\,(x) \cap F \neq \emptyset \implies x \in F]\,,$$

is equivalent to saying F is closed in X. This notion gives us a nice way to judge when a set F is **NOT** closed in X. For that to happen there must be an $x \in X$ such that no matter any radius r we take, the ball of radius r about x always intersects F but x does not belong to F.

Example 1.1.2. *We show a few examples of open and closed sets in a metric space (X, d).*

(1) *We observe that $X, \emptyset$ are both open and closed sets. In fact it is clear that X is open and that $\emptyset$ is open because it has no points. Then taking complements we see that they both are closed.*

(2) *The open ball is an open set. To see this let $G = \mathcal{B}_r\,(x)$ be an open ball in X. To show that it is an open set we need, for any point y in G, to find a radius $\delta > 0$ such that the ball $\mathcal{B}_\delta\,(y) \subset G$. Let $\delta = r - d(y, x)$. This is greater than 0 since y belongs to G. Now*

to show that $\mathcal{B}_\delta(y) \subset G$ we need to show that for any $z \in \mathcal{B}_\delta(y)$, z must be in G i.e. $d(z,x) < r$. But this follows from the following argument,

$$\begin{aligned} d(z,x) &\leq d(z,y) + d(y,x), \\ &\leq \delta + r - \delta, \\ &= r. \end{aligned}$$

(3) *The closed ball is a closed set. To see this let $F = \overline{\mathcal{B}_r(x)}$ be a closed ball in X. To show that F is closed we need to show that for any $y \in X$ if $y \notin F$ then we must be able to find a radius $\delta > 0$ such that $\mathcal{B}_\delta(y)$ doesn't intersect F. Let y be a point in $X - F$. Hence $d(y,x) > r$. Let $\delta = d(y,x) - r$. Hence $\delta > 0$. For any $z \in \mathcal{B}_\delta(y)$,*

$$\begin{aligned} d(x,z) + d(z,y) \geq d(x,y) & \\ &\implies d(x,z) \geq d(x,y) - d(z,y), \\ &\implies d(x,z) \geq r. \end{aligned}$$

Hence $\mathcal{B}_\delta(y) \cap F = \emptyset$.

(4) *Any finite subset of X is closed. In fact if $F = \{x_1, x_2, x_3, \ldots, x_n\}$, then for any $y \in X - F$, let $\delta = \min\{d(y,x_i) \ : \ 1 \leq i \leq n\}$. Then $\mathcal{B}_{\frac{\delta}{2}}(y) \cap F = \emptyset$.*

Open sets are closed under arbitrary unions and closed sets are closed under arbitrary intersections. We make this precise in the next proposition.

Proposition 1.1.1. *Let (X,d) be a metric space and let $\mathscr{G}$ be the collection of all open sets in X. Then,*

(1) *$X, \emptyset$ are in $\mathscr{G}$.*

(2) *If $G_1, G_2, \ldots, G_n$ are n open sets in $\mathscr{G}$, then $\bigcap_{i=1}^{n} G_i$ is also in $\mathscr{G}$.*

(3) *If $\{G_\alpha \in \mathscr{G} \ : \ \alpha \in I\}$ is an arbitrary sub collection of open sets in $\mathscr{G}$ then $\bigcup_{\alpha \in I} G_\alpha$ is also in $\mathscr{G}$.*

Proof. We prove in order.

(1) We showed this in the example.

(2) Let x be in $\bigcap_{i=1}^{n} G_i$. Then x is in each G_i and since G_i is open there is an $r_i > 0$ such that $\mathcal{B}_{r_i}(x) \subset G_i$. Take $r = \min\{r_i \ : \ 1 \leq i \leq n\}$. Then $r > 0$ and $\mathcal{B}_{\frac{r}{2}}(x) \subset G_i$ for all i's. Hence

$$\mathcal{B}_{\frac{r}{2}}(x) \subset \bigcap_{i=1}^{n} G_i.$$

(3) Let x be in $\bigcup_{\alpha \in I} G_\alpha$. Then there is an α in I such that x is in G_α. Since G_α is open there is an $r > 0$ such that $\mathcal{B}_r(x) \subset G_\alpha$. But $G_\alpha \subset \bigcup_{\alpha \in I} G_\alpha$ and hence

$$\mathcal{B}_r(x) \subset \bigcup_{\alpha \in I} G_\alpha.$$

□

Proposition 1.1.2. *Let (X, d) be a metric space and let $\mathscr{F}$ be the collection of all closed sets in X. Then,*

(1) $X, \emptyset$ are in $\mathscr{F}$.

(2) If $F_1, F_2, \dots, F_n$ are n closed sets in $\mathscr{F}$, then $\bigcup_{i=1}^{n} F_i$ is also in $\mathscr{F}$.

(3) If $\{F_\alpha \in \mathscr{F} \ : \ \alpha \in I\}$ is an arbitrary sub collection of closed sets in $\mathscr{F}$ then $\bigcap_{\alpha \in I} F_\alpha$ is also in $\mathscr{F}$.

Proof. De Morgan's law of set complements. □

One must be very careful about the universe when talking about open and closed sets in a metric space. For example if we consider $X = \mathbb{R}^2$ with the euclidean metric and $\mathbb{R} = Y \subset \mathbb{R}^2$ with the same Euclidean metric, then an interval (a, b) is open in Y but is **NOT** open in X. To make this precise we will define the notion of relative openess (or closedness).

Definition 1.1.6 (Open balls in subspace)**.** *Let (X, d) be a metric space and consider $Y \subset X$. We define an open ball in (Y, d) of radius r about a point $y \in Y$ as the set,*

$$\mathcal{B}_r^Y(y) = \{z \in Y \ : \ d(z, y) < r\} = \mathcal{B}_r(y) \cap Y.$$

Example 1.1.3. *Let $X = \mathbb{R}$ and let $Y = [0, 1]$ with the metric given by the absolute value. Then an open ball in Y of radius $\frac{1}{2}$ about 0 is $\mathcal{B}_{\frac{1}{2}}^Y(0) = [0, \frac{1}{2})$ where as the open ball in X is $(-\frac{1}{2}, \frac{1}{2})$.*

Proposition 1.1.3. *Let (X, d) be a metric space and let Y be a subset of X.*

(1) A subset G of Y is relatively open in Y if and only if there is an open subset U in X with $G = U \cap Y$.

(2) A subset F of Y is relatively closed in Y if and only if there is a closed subset U in X with $F = U \cap Y$.

Proof. We will show the proof for only (1).

Let $G \subset Y$ be relatively open in Y. This means that for any $x \in G$ there is radius r_x such that an open ball in Y, $\mathcal{B}^Y_{r_x}(x) \subset G$. Let $U = \bigcup_{x \in G} \mathcal{B}_{r_x}(x)$. Then by 1.1.1, U is open in X. We will show that $G \subset U \cap Y$ and $U \cap Y \subset G$.

If $z \in G$ then $z \in \mathcal{B}^Y_{r_z}(z) = \mathcal{B}_{r_z}(z) \cap Y \subset U \cap Y$. Thus $G \subset U \cap Y$. If $z \in U \cap Y$. Then there is an $x \in G$ such that $z \in \mathcal{B}_{r_x}(x)$ and $z \in Y$ that is $z \in \mathcal{B}^Y_{r_x}(x)$ which is a subset of G. Thus $z \in G$ i.e. $U \cap Y \subset G$.

Let U be an open set in X such that $G = U \cap Y$, that is to say,

$$G \subset U \cap Y \subset G.$$

If $x \in G$ then $x \in U \cap Y$ and since U is open there is a $r > 0$ such that $\mathcal{B}_r(x) \subset U$. That means that

$$\mathcal{B}^Y_r(x) = \mathcal{B}_r(x) \cap Y \subset U \cap Y \subset G.$$

□

Open and closed sets enable us to look at points in a metric space with a geometric lens. Given a set, a point may either be *inside* it, *outside* it or on the *edge*. These notions can be made precise.

Definition 1.1.7 (Interior of a set)**.** *Let (X, d) be a metric space and let A be a subset of X. The interior of a set A, denoted by A° is the set defined by,*

$$A^\circ = \bigcup \{G \subset X \ : \ G \text{ is open and } G \subset A\}.$$

Thus if $x \in A^\circ$ then there is an $r > 0$ such that the ball $\mathcal{B}_r(x)$ is entirely contained in A.

Definition 1.1.8 (Closure of a set)**.** *Let (X, d) be a metric space and let A be a subset of X. The closure of A, denoted by $\operatorname{cl} A$, is the set defined by,*

$$\operatorname{cl} A = \bigcap \{F \subset X \ : \ F \text{ is closed and } F \supset A\}.$$

Thus if $x \in \operatorname{cl} A$ then any closed set F containing A must include the point x.

Definition 1.1.9 (Boundary of a set)**.** *Let (X, d) be a metric space and let A be a subset of X. The boundary of A, denoted by ∂A, is the set defined by*

$$\partial A = \operatorname{cl} A \cap \operatorname{cl}(X - A).$$

Remark 1.1.1. *Since $\emptyset$ is open and is contained in every set, the interior of any set always contains the emptyset. This however, may be all the interior of a set, in other words there could be a $A \subset X$ such that $A^\circ = \emptyset$. Since X is closed and contains any set, it may happen that it is the only closed set containing the set in other words there could be a $A \subset X$ such that $\operatorname{cl} A = X$.*

It follows from 1.1.1 that A° is open (being the union of open sets) and from 1.1.2 that the $\operatorname{cl} A$ *is closed (being the intersection of closed sets).*

The following Proposition gives a useful characterization of the interior and closure of sets.

Proposition 1.1.4. *Let (X, d) be a metric space and let $A \subset X$. Then,*

(1) $x \in A^\circ$ if and only if there is an $r > 0$ such that $\mathcal{B}_r(x) \subset A$.

(2) $x \in \operatorname{cl} A$ if and only if for every $r > 0$ the ball $\mathcal{B}_r(x) \cap A \neq \emptyset$.

Proof. We prove in order.

- If $x \in A^\circ$ then it follows from the definition that it belongs to an open set $G \subset X$ and hence there is an $r > 0$ such that $\mathcal{B}_r(x) \subset G \subset A$.

  Assume there is an $r > 0$ such that the $\mathcal{B}_r(x) \subset A$. Since an open ball is an open set, let $G = \mathcal{B}_r(x)$. Then $x \in G \subset A$ and hence $x \in A^\circ$.
- Take an $x \in \operatorname{cl} A$. Fix an $r > 0$. Assume $\mathcal{B}_r(x) \cap A = \emptyset$ i.e. $\mathcal{B}_r(x) \subset (X - A)$. This means that $(X - \mathcal{B}_r(x)) \supset A$. Since $\mathcal{B}_r(x)$ is open, $(X - \mathcal{B}_r(x))$ is a closed set containing A and hence must contain $\operatorname{cl} A$. Thus $\mathcal{B}_r(x) \cap \operatorname{cl} A = \emptyset$. Hence we reach a contradiction (in assuming $\mathcal{B}_r(x) \cap A = \emptyset$) since $x \in \operatorname{cl} A$ and $\operatorname{cl} A$ is closed.

  To show the other implication we prove the contrapositive. Assume $x \notin \operatorname{cl} A$ i.e. $x \in X - \operatorname{cl} A$. Since $\operatorname{cl} A$ is closed, this means that there is an $r > 0$ such that $\mathcal{B}_r(x) \subset X - \operatorname{cl} A$. But $(X - \operatorname{cl} A) \subset (X - A)$ since $A \subset \operatorname{cl} A$. Hence $\mathcal{B}_r(x) \cap A = \emptyset$.

$\square$

Example 1.1.4. *Let X be $\mathbb{R}$ and $A = \mathbb{Q}$. We know from elementary analysis that any interval contains a rational number and any interval with rational endpoints contain a real number. Thus $\operatorname{cl} \mathbb{Q} = \mathbb{R}$. Similarly any interval with rational endpoints contain irrational number and so $\mathbb{Q}^\circ = \emptyset$. Using the same reasoning we see that $\operatorname{cl}(\mathbb{R} - \mathbb{Q}) = \mathbb{R}$ and $(\mathbb{R} - \mathbb{Q})^\circ = \emptyset$. Thus $\partial \mathbb{Q} = \mathbb{R}$.*

Example 1.1.5. *We will show that the closed ball is not generally the closure of the open ball. Let (X, d) be the discrete metric. Then for any $x \in X$, $\mathcal{B}_r(x) = \{x\} = \operatorname{cl} \mathcal{B}_r(x)$. But $\overline{\mathcal{B}_r(x)} = X$.*

Proposition 1.1.5. *Let (X, d) be a metric space and let A be a subset of X. Then,*

(1) A is closed if and only if $A = \operatorname{cl} A$.

(2) A is open if an only if $A = A^\circ$.

(3) $\operatorname{cl} A = X - (X - A)^\circ$.

(4) $A^\circ = X - \operatorname{cl}(X - A)$.

(5) $\partial A = \operatorname{cl} A - A^\circ$.

(6) If $A_1, \ldots, A_n$ *are subsets of* X*, then* $\operatorname{cl}(\bigcup_{k=1}^{n} A_k) = \bigcup_{k=1}^{n} \operatorname{cl} A_k$.

Proof. We prove in order.

- First note that $A \subset \operatorname{cl} A$ for any $A \subset X$. If A is closed then since $A \subset A$ and $\operatorname{cl} A \subset F$ for all F closed and containing A, $\operatorname{cl} A \subset A$. Thus if A is closed then $A = \operatorname{cl} A$.

  Since $\operatorname{cl} A$ is closed and $A = \operatorname{cl} A$, A is closed.

- Note that A° is open and $A^\circ \subset A$. Let A be open and let x be in A. Hence there is an $r > 0$ such that $\mathcal{B}_r(x) \subset A$. But by 1.1.4, x is in A° and hence $A \subset A^\circ$.

- For any set $B \subset X$, let us denote by B^c the set $X - B$. With this terminology, first observe that $(A^\mathsf{c})^\circ \subset A^\mathsf{c}$ and hence, $A \subset ((A^\mathsf{c})^\circ)^\mathsf{c}$. Since $((A^\mathsf{c})^\circ)^\mathsf{c}$ is closed and contains A, we must have $\overline{A} \subset ((A^\mathsf{c})^\circ)^\mathsf{c}$.

  Let x be an element of $((A^\mathsf{c})^\circ)^\mathsf{c}$ and hence $x \notin (A^\mathsf{c})^\circ$, which means that for any $r > 0$ $\mathcal{B}_r(x) \not\subset A^\mathsf{c}$; which is equivalent to saying that $\mathcal{B}_r(x) \cap A \neq \emptyset$ for any $r > 0$ and hence $x \in \overline{A}$. Thus, we have shown that $((A^\mathsf{c})^\circ)^\mathsf{c} \subset A$.

- Again, using the notation from above, note that $\operatorname{cl}(A^\mathsf{c}) \supset A^\mathsf{c}$ and hence $(\operatorname{cl}(A^\mathsf{c}))^\mathsf{c} \subset A$. Since $(\operatorname{cl}(A^\mathsf{c}))^\mathsf{c}$ is an open set contained in A, we must have $(\operatorname{cl}(A^\mathsf{c}))^\mathsf{c} \subset A^\circ$.

  Let x be an element of A°. Then there is an $r > 0$ such that $\mathcal{B}_r(x) \subset A$, which means that $\mathcal{B}_r(x) \cap A^\mathsf{c} = \emptyset$. But this means that $x \notin \operatorname{cl}(A^\mathsf{c})$, which is equivalent to saying that $x \in (\operatorname{cl}(A^\mathsf{c}))^\mathsf{c}$. Hence, we have shown that $A^\circ \subset (\operatorname{cl}(A^\mathsf{c}))^\mathsf{c}$.

- Using the notation from above, note that $\partial A = \operatorname{cl} A \cap \operatorname{cl} A^\mathsf{c}$. From (4), we see that $\operatorname{cl} A^\mathsf{c} = (A^\circ)^\mathsf{c}$ and hence $\partial A = \operatorname{cl} A \cap (A^\circ)^\mathsf{c} = \operatorname{cl} A - A^\circ$.

- Let $A = \bigcup_{k=1}^{n} A_k$. Note that $A \supset A_k$ and hence $\operatorname{cl} A \supset \operatorname{cl} A_k$ for each k. Thus $\operatorname{cl} A \supset \bigcup_{k=1}^{n} \operatorname{cl} A_k$. Each $\operatorname{cl} A_k$ is closed and since **finite** union of closed sets are close $\bigcup_{k=1}^{n} \operatorname{cl} A_k$ is a closed set. Moreover since $A_k \subset \operatorname{cl} A_k$, we have $A \subset \bigcup_{k=1}^{n} \operatorname{cl} A_k$. Thus we have a closed set containing A and hence $\operatorname{cl} A \subset \bigcup_{k=1}^{n} \operatorname{cl} A_k$.

□

Remark 1.1.2. *Sometimes when it is obvious from the context that the set difference are being taken w.r.t. X, we usually denote $X - A$ as A^{c}. With such a notation (3), (4) in the theorem above can be written as $(A^{\mathsf{c}})^\circ = (\operatorname{cl} A)^{\mathsf{c}}$ and $\operatorname{cl} A^{\mathsf{c}} = (A^\circ)^{\mathsf{c}}$.*

A very useful concept involving the closure is the notion of density.

Definition 1.1.10 (Dense subset)**.** *A subset E of a metric space (X, d) is dense if $\operatorname{cl} E = X$. This means that for any $x \in X$, $\mathcal{B}_r(x) \cap E \neq \emptyset$ for any radius r.*

Definition 1.1.11 (Separable)**.** *A metric space (X, d) is separable if it has a countable dense subset.*

Example 1.1.6. *We show some examples of dense subset.*

- *The rational numbers are dense in $\mathbb{R}$. This is because any interval around a real number must contain rational numbers. Thus $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$. This means that $\mathbb{R}^n$ is separable. This follows from a general fact that if A_1, A_2 are dense subsets of X_1, X_2, then $A_1 \times A_2$ is dense in $X_1 \times X_2$. We will show this fact when we define a product metric space.*
- *If X is an set and d is the discrete metric, then the only dense subset of X is X itself.*

We now look into the important topic of sequences and its convergence in a metric space.

1.2. Sequences and completeness

Let (X, d) be a metric space.

Definition 1.2.1 (Sequence)**.** *A sequence (x_n) in X is a function that maps the positive integers to points in the metrics space X i.e. $n \mapsto x_n$. A subsequence of a sequence is defined by (x_{n_k}), is a composition of the sequence function with a subsequence index non-decreasing function that maps positive integers to positive integers i.e. $k \mapsto n_k$ and $k_1 \leq k_2$ implies $n_{k_1} \leq n_{k_2}$. Hence a subsequence is the function that maps positive integers k to points in X i.e. $k \mapsto x_{n_k}$.*

Example 1.2.1. *Let (R, d) be the standard metric space in $\mathbb{R}$. Then (x_n) given by $1, -1, 1, -1, \cdots$ is a sequence, while (x_{n_k}) given by $1, -1, 1, 1, 1\cdots$ is a subsequence. We can explicitly write the subsequence index function as $1 \mapsto 1$, $2 \mapsto 2$, $3 \mapsto 3$, $4 \mapsto 5$, $5 \mapsto 7$ and so on.*

Definition 1.2.2 (Convergence)**.** *A sequence (x_n) is said to converge in X if for every $\epsilon > 0$ there is an integer N such that $d(x, x_n) < \epsilon$, whenever $n \geq N$. We denote it by $x_n \to x$ or $x = \lim_{n\to\infty} x_n$. Logically this means,*

$$\forall\,(\epsilon > 0)\,\exists\left(N \in \mathbb{Z}^+\right)\forall\left(n \in \mathbb{Z}^+\right)\left[n \geq N \implies d(x_n, x) < \epsilon\right].$$

What does it mean for a sequence to not converge to x? From our logical implication, this means that there is an $\epsilon > 0$ such that no matter what N we take, we can find a n in the sequence with $n \geq N$ but $d(x_n, x) \geq \epsilon$.

Proposition 1.2.1. *If $x_n \to x$ then any subsequence x_{n_k} also coverges to x.*

Proof. Let $\epsilon > 0$ be given. Since (x_n) is convergent, there is an N such that for all $n \geq N$, $d(x_n, x) < \epsilon$. However, when $n \geq N$, then $n_k \geq N$ since for all k, $n_k \geq n$ (can be shown by induction). Thus $d(x_{n_k}, x) < \epsilon$. $\square$

There is a deep connection between closed sets in metric spaces and convergence of sequences.

Proposition 1.2.2. *A subset $F \subset X$ is closed if and only if whenever (x_n) is a sequence in F and $x_n \to x$, it follows that $x \in F$.*

Proof. Let F be closed. Assume there is a sequence (x_n) in F such that $x_n \to x$. If $x \notin F$, then by the definiton of closed set there is a radius $r > 0$ such that $\mathcal{B}_r(x) \cap = \emptyset$. But since $x_n \to x$ there is an N such that $d(x_n, x) < r$ for all $n \geq N$. Hence x must belong to F.

Let x be an element of X. Assume $\mathcal{B}_r(x) \cap F \neq \emptyset$ for any $r > 0$. In particular, for $r = 1$, there is an x_1 such that $d(x_1, x) < 1$. Going on inductively we have for each n a x_n such that $d(x_n, x) < 1/n$. This means that we constructed a sequence (x_n) such that $x_n \to x$. By our hypothesis x belongs to F and hence F is closed. (See the second logical implication following the definition of closed set.) $\square$

Remark 1.2.1. *In the proof above, we uncovered a nice observation about the closure of a set A. If x is in the closure of A then for all $r > 0$, $\mathcal{B}_r(x) \cap A \neq \emptyset$. We saw, how this leads to a sequence (x_n) in A such that $x_n \to x$. However, this sequence may have only a finite number of distinct terms. In other words we could get a **constant** sequence $x_1 = x_2 = \cdots = x$.*

Definition 1.2.3 (Limit point). *If $A \subset X$, then a point $x \in X$ is called a limit point of A if for every $\epsilon > 0$ there is a point a in $\mathcal{B}_\epsilon(x) \cap A$ with $a \neq x$. Logically this means,*

$$\forall (\epsilon > 0) \, \exists (a \in A) \, [a \neq x \text{ and } d(a, x) < \epsilon] .$$

What does it mean for a point x to be **NOT** a limit point of A. If that is the case then there must be an ϵ such that no matter what point a in A we take, if $a \neq x$ then the distance between a and x is larger than ϵ.

Definition 1.2.4 (Isolated point). *A point $x \in A$ that is not a limit point of A is called an isolated point. Thus if x is an isloated point of A, then there must be an $\epsilon > 0$ such that $\mathcal{B}_r(x) \cap (A - \{x\}) = \emptyset$.*

Example 1.2.2. *The following are some examples of limit points/isolated points.*

- *Let $X = \mathbb{R}$ and let $A = (0, 1) \cup 2$. Note that 2 is an isolated point of A because the ball $\mathcal{B}_{\frac{1}{3}}(2)$ doesn't intersect $(0, 1)$. Each point in $[0, 1]$ is a limit point of A.*
- *If $X = \mathbb{R}$ and $A = \mathbb{Q}$, then every point of X is a limit point of A and A has no isolated points.*
- *If $X = \mathbb{R}$ and $A = \left\{\frac{1}{n} \,:\, n \in \mathbb{Z}^+\right\}$, then 0 is the limit point of A, while all points of A are isolated points of A.*

Proposition 1.2.3. *Let A be a subset of X. Then,*

(1) A point x is a limit point of A if and only if there is a sequence of distinct points in A that converges to x.

(2) $\operatorname{cl} A = A \cup \{x \,:\, x \text{ is a limit point of } A\}$.

(3) A is closed in X if and only if it contains all its limit points.

Proof. We prove in order.

(1) Let x be a limit point of A. Then for every $\epsilon > 0$ there is an $a \in A$ such that $a \neq x$ and $d(a, x) < \epsilon$. We will define a sequence inductively. First pick $\epsilon_1 = 1$. Then there is an $a_1 \neq x$ such that $d(a_1, x) < \epsilon_1$. Now take radius to be $\frac{1}{2}$. Is there an $a_2 \neq a_1 \neq x$ such that $d(a_2, x) < \frac{1}{2}$? It could so happen that the a_1 we found actually satisfies $d(a_1, x) < \frac{1}{2}$. Thus to remedy this let $\epsilon_2 = \min\left\{\frac{1}{2}, d(a_1, x)\right\}$. Thus there exists an $a_2 \neq a_1$ such that $d(a_2, x) < \epsilon_2$. Inductively, we have found for each n there is an a_n different from $a_1, a_2, \ldots, a_{n-1}$ and $d(a_n, x) < \epsilon_n \leq \frac{1}{n}$. Taking n large enough we get a sequence of distinct terms (a_n) such that $a_n \to x$.

Let (a_n) be a sequence of distinct points in A such that $a_n \to x$. Note that $x \neq a_n$ for any n. For any $\epsilon > 0$ there is an N such that $d(a_n, x) < \epsilon$. Thus for any ϵ there is an a_n such that $a_n \neq x$ and $d(a_n, x) < \epsilon$. Hence x is a limit point of A.

(2) If x is a point of A then x certainly belongs to $\operatorname{cl} A$. Let x be a limit point of A. Then there is a sequence of distinct points (a_n) in A and hence in $\operatorname{cl} A$ such that $a_n \to x$. But since $\operatorname{cl} A$ is closed, x must be in $\operatorname{cl} A$. Thus, $A \cup \{x \;:\; x \text{ is a limit point of } A\} \subset \operatorname{cl} A$.

Let x be a point in $\operatorname{cl} A$ and assume $x \notin A$. Then for any $\epsilon > 0$, the $\mathcal{B}_\epsilon(x) \cap A \neq \emptyset$. Since $x \notin A$, there must be an $a \neq x$ such that a is in A and $d(x, a) < \epsilon$. Hence x is a limit point of A. Thus, $\operatorname{cl} A \subset A \cup \{x \;:\; x \text{ is a limit point of } A\}$.

(3) If A is closed then $A = \operatorname{cl} A$ and hence from the result above it must contain all its limit points.

□

The closure of a set can also be characterized in a different way that has a geometric motivation. For this we define the distance between a point and a set in a metric space.

Definition 1.2.5 (Distance)**.** *If $A \subset X$ and $x \in X$, then the distance from x to A is the number, denoted as* $\operatorname{dist}(x, A)$, *given by,*

$$\operatorname{dist}(x, A) = \inf \{d(a, x) \;:\; a \in A\}.$$

When x is in A, then clearly $\operatorname{dist}(x, A) = 0$. Is it possible for $\operatorname{dist}(x, A) = 0$ when $x \notin A$?

Proposition 1.2.4. *If $A \subset X$, then* $\operatorname{cl} A = \{x \in X \;:\; \operatorname{dist}(x, A) = 0\}$.

Proof. Let x be in $\operatorname{cl} A$. Then there is a sequence (x_n) in A such that $x_n \to x$. This means that for any ϵ there is an N such that $d(x_n, x) < \epsilon$ whenever $n \geq N$. Hence $\inf \{d(a, x) \;:\; a \in A\} < \epsilon$. Since ϵ was arbitrary, $\operatorname{dist}(x, A) = 0$.

Let x be a point in X such that $\operatorname{dist}(x, A) = 0$. Then, for any $\epsilon > 0$ there is an $a \in A$ such that $d(a, x) < \epsilon$. This means that for any $\epsilon > 0$ $\mathcal{B}_\epsilon(x) \cap A \neq \emptyset$. Hence $x \in \operatorname{cl} A$. □

We need to know apriori the limit of a sequence to check if we have convergence. This notion is not very helpful as we would like to know if a given sequence convergence without any knowledge of its limit. We can do this for certain sequences called the Cauchy sequence.

Definition 1.2.6 (Cauchy sequence)**.** *A sequence (x_n) in X is a Cauchy sequence if,*

$$\lim_{n,m\to\infty} d(x_n, x_m) = 0.$$

This means that for any $\epsilon > 0$ there is an $N \in \mathbb{Z}^+$ such that $d(x_n, x_m) < \epsilon$ whenever $n, m \geq N$.

Definition 1.2.7 (Complete metric space)**.** *A metric space X is complete if every Cauchy sequence converges.*

Remark 1.2.2. *A few observations can be made about Cauchy sequences.*

- *A convergent sequence is Cauchy.*
- *A Cauchy sequence with a convergent subsequence is convergent.*

A useful geometric notion can be given to any set in a metric space.

Definition 1.2.8 (Diameter)**.** *For any set $E \subset X$, we define its diameter as*

$$\operatorname{diam} E = \sup\{d(x,y) \;:\; x, y \in E\}.$$

Proposition 1.2.5. *For any set $E \subset X$, $\operatorname{diam} E = \operatorname{diam} \operatorname{cl} E$.*

Proof. For any p, q in E, p, q are in $\operatorname{cl} E$ and hence $\operatorname{diam} E \leq \operatorname{diam} \operatorname{cl} E$. Let x, y be points in $\operatorname{cl} E$ such that $x \neq y$. Then ther are point p, q in E such that $d(p,x) < \frac{\epsilon}{2}$ and $d(q,y) \leq \frac{\epsilon}{2}$. By (repeated use of) triangle inequlality,

$$d(x,y) < d(x,p) + d(p,q) + d(q,y) < d(p,q) + \epsilon.$$

This means that $\operatorname{diam} \operatorname{cl} E \leq \operatorname{diam} E + \epsilon$. Since ϵ was arbitray we get $\operatorname{diam} \operatorname{cl} E \leq \operatorname{diam} E$. □

Theorem 1.2.1 (Cantor Intersection theorem)**.** A metric space is complete if and only if (F_n) is a sequence of non-empty subsets satisfying,

(1) Each F_n is closed,
(2) $F_1 \supset F_2 \supset F_3 \ldots,$
(3) $\operatorname{diam} F_n \to 0,$

then $\bigcap_{n=1}^{\infty} F_n$ is a single point.

Proof. Assume we have a complete metric space (X, d) and let (F_n) be a sequence of arbitrary closed sets in X such that,

(1) Each F_n is closed,
(2) $F_1 \supset F_2 \supset F_3 \ldots,$
(3) $\operatorname{diam} F_n \to 0,$

Let (x_n) be a sequence of points in X such that $x_i \in F_i$.

Claim: (x_n) is Cauchy. Fix an $\epsilon > 0$. Since $\operatorname{diam} F_n \to 0$, there is an N such that $\operatorname{diam} F_n < \epsilon$ whenever $n \geq N$. For any $n, m \geq N$, $x_n, x_m \in F_N$ since $F_N \supset F_{N+1} \dots$ and hence $d(x_n, x_m) < \epsilon$.

Since (X, d) is complete, there is an $x \in X$ such that $x_n \to x$. Note that (x_n) is in F_1 and since F_1 is closed x is in F_1. But the sequence $(x_{n>1})$ is in F_2 and since F_2 is closed x is in F_2. Thus continuing this way, since each F_n is closed, $x \in \bigcap_{n=1}^{\infty} F_n$. Hence $\bigcap_{n=1}^{\infty} F_n$ is not empty. If y is in $\bigcap_{n=1}^{\infty} F_n$, then $d(x, y) < \operatorname{diam} F_n$ for each N and hence $d(x, y) = 0$ i.e $x = y$.

Now assume we have a sequence (F_n) of arbitrary closed sets in X such that IF

(1) Each F_n is closed,

(2) $F_1 \supset F_2 \supset F_3 \dots$,

(3) $\operatorname{diam} F_n \to 0$,

is true then $\bigcap_{n=1}^{\infty} F_n$ is a single point. Let (x_n) be a Cauchy sequence in X. Let $F_k = \operatorname{cl}\{x_n : n \geq k\}$. Note that $F_1 \supset F_2 \supset F_3 \dots$ and each F_n is closed.

Claim: $\operatorname{diam} F_n \to 0$. Fix an $\epsilon > 0$. Then there is an N such that $d(x_n, x_m) < \epsilon$ whenever $n \geq N$. This means that $\operatorname{diam} F_N \epsilon$. But $F_N \supset F_{N+1} \dots$ and hence $\operatorname{diam} F_n \leq \operatorname{diam} F_N$ for every $n \geq N$ i.e $\operatorname{diam} F_n < \epsilon$ whenever $n \geq N$. Thus $\operatorname{diam} F_n \to 0$.

Hence $\bigcap_{n=1}^{\infty} F_n$ contains only one point, lets call it x. For any x_n, $d(x, x_n) < \operatorname{diam} F_n$ which goes to 0 as $n \to \infty$. □

Example 1.2.3. $\mathbb{R}, \mathbb{R}^n$ *are complete spaces. The proof is part of an elementary real analysis course.*

Proposition 1.2.6. *If (X, d) is a complete metric space and $Y \subset X$, then (Y, d) is complete if and only if Y is closed in X.*

Proof. Assume (Y, d) is complete. Let (x_n) be a sequence in Y such that $x_n \to x$ for some x in X. Since a convergent sequence is Cauchy, (x_n) is Cauchy and since Y is complete there is a $y \in Y$ such that $x_n \to y$. But since limits are unique $x = y$ and hence Y is closed.

Assume Y be closed in X. Let (x_n) be a Cauchy sequence. Since $Y \subset X$, (x_n) is cauchy in X. Since X is complete, there is a $x \in X$ such that $x_n \to x$. But since Y is closed x is in Y and hence (Y, d) is complete. □

Definition 1.2.9 (Bounded set)**.** *A subset A in a metric space (X, d) is bounded if* $\operatorname{diam} A < \infty$.

Proposition 1.2.7. *We observe the following for bounded sets.*

(1) A subset A of (X, d) is bounded if and only if for any $x \in X$, there is an $r > 0$ such that $A \subset \mathcal{B}_r(x)$.

(2) The union of a finite number of bounded sets is bounded.

(3) A cauchy sequence in (X, d) is a bounded set.

Proof. We prove in order.

(1) Let A be bounded. Thus $\operatorname{diam} A < \infty$ i.e. there is an $M \in \mathbb{Z}^+$ such that $\operatorname{diam} A < M$. Thus, $d(p, q) < M$ for any $p, q \in A$. Let x be any point in X. Fix a point $p_0 \in A$ and let $r > d(x, p_0) + M$. Then, for any $p \in A$,

$$\begin{aligned} d(x, q) &\leq d(x, p_0) + d(p_0, q) \\ &< r. \end{aligned}$$

Assume for any x, there is an r such that $A \subset \mathcal{B}_r(x)$. Let p_0 be an element of A (assume A is not empyt). Then there is an $r > 0$ such that $A \subset \mathcal{B}_r(x)$. Thus, for any $p, q \in A$, $d(p, q) < 2r$. Hence, $\operatorname{diam} A \leq 2r < \infty$.

(2) Let $\{A_i : 1 \leq i \leq n\}$ be a finite collection of bounded sets. Thus there are r_i such that $A_i \subset \mathcal{B}_{r_i}(x)$ for each $1 \leq i \leq n$ for any $x \in X$. Let $r > \sum_{i=1}^{n} r_i$. Then for each i, $A_i \subset \mathcal{B}_r(x)$ and hence, $\bigcup_{i=1}^{n} A_i \subset \mathcal{B}_r(x)$.

(3) Let (x_n) be a cauchy sequence and let $A = \{x_n : n \in \mathbb{Z}^+\}$. Fix $\epsilon = 1$. Then, there is an N such that $d(x_n, x_m) < 1$ for all $n, m \geq N$. Let $B_N = \{x_n : n \geq N\}$. Thus, $\operatorname{diam} B_N \leq 1$ and hence B_N is bounded. Let $A_N = A - B_N$. Then, A_N is a finite set and is bounded (diameter of a finite set is always the maximum distance between its points.) Hence $A = A_N \cup B_N$ is bounded.

$\square$

1.3. Continuity

We will extended the notion of continuous functions to a mapping between two metric spaces.

Definition 1.3.1 (Continuous function)**.** *If (X, d) and (Y, ρ) are two metric spaces, a function $f\ : X \to Y$ is continuous at a point $a \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that when $d(x, a) < \delta$, then $\rho(f(x), f(a)) < \epsilon$. We say that f is continuous on X if f is continuous at every point $a \in X$.*

We can turn the definition of continuity relating to sequences in the following way:

Proposition 1.3.1. *If (X, d) and (Y, ρ) are metric spaces and $f\ : X \to Z$ is a function, then f is continuous if and only if whenever (x_n) is a sequence in X such that $x_n \to x$ then, $f(x_n) \to f(a)$.*

Proof. Suppose f is continuous. Fix an $\epsilon > 0$. Then, there is a $\delta > 0$ such that IF $d(x, a) < \delta$, then $\rho(f(x), f(a)) < \epsilon$. For this δ, there is an N such that $d(x_n, a) < \delta$ for all $n \geq N$. Hence, $\rho(f(x_n), f(a)) < \epsilon$. This means that $f(x_n) \to f(a)$.

Assume f is not continuous. Then there is an $\epsilon > 0$ such that for any $\delta > 0$, $d(x, a) < \delta$ but $\rho(f(x), f(a)) \geq \epsilon$. Pick $\delta = 1$. There is a $x_1 \in X$ such that $d(x_1, a) < 1$ and $\rho(f(x_1), f(a)) \geq \epsilon$. Continuing this way, for every $\delta_n = \frac{1}{n}$, there is an $x_n \in X$ such that $d(x_n, a) < \frac{1}{n}$ but $\rho(f(x_n), f(a)) \geq \epsilon$. Hence, $x_n \to x$ but $f(x_n) \not\to f(a)$. $\square$

The Proposition above can be used to prove the following proposition which tells us that continous functions work well with algebraic operations.

Proposition 1.3.2. *If (X, d) is a metric space and f, g are continuous functions from X into $\mathbb{R}$, then*

(1) $(f \pm g)$ is continuous. Here $(f \pm g)(x)$ is defined to be $f(x) \pm g(x)$ for any $x \in X$.
(2) (fg) is continuous. Here $(fg)(x)$ is defined to be $f(x)g(x)$ for any $x \in X$.
(3) If $f(x)$ is not 0 for all $x \in X$, then $\frac{1}{f}$ is a continuous function.

We now want to look at functions that are continuous on the entire metric space. In that case, we can look at an equivalent notion of continuity. Note that when $d(x, a) < \delta$ then $\rho(f(x), f(a)) < \epsilon$ means that $\mathcal{B}_\delta(a)$ is getting mapped inside the $\mathcal{B}_\epsilon(f(a))$. Thus, we can restate the definition of continuity as follows,

Proposition 1.3.3. *If (X, d) and (Y, ρ) are two metric spaces, a function $f\ : X \to Y$ is continuous at a point $a \in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that $f(\mathcal{B}_\delta(a)) \subset \mathcal{B}_\epsilon(f(a))$. We say that f is continuous on X if f is continuous at every point $a \in X$.*

Theorem 1.3.1. Let (X, d) and (Y, ρ) be two metric spaces and consider $f : X \to Y$. Then, the following are equivalent.

(1) f is continuous function on X.
(2) If G is an open subset of Y, then $f^{-1}(G)$ is an open subset of X.
(3) If F is a closed subset of Y, then $f^{-1}(F)$ is a closed subset of X.

Proof. We will show that (1) $\iff$ (2). First supposet that f is continuous on X and let G be an open subset of Y. We need to show that for any $p \in f^{-1}(G)$ there is an $r > 0$ such that $\mathcal{B}_r(p) \subset f^{-1}(G)$. Let p be any point in $f^{-1}(G)$. Thus, $f(p)$ is a point in G. Since G is open, there is an $\epsilon > 0$ such that $\mathcal{B}_\epsilon(f(p)) \subset G$. Since f is continuous at p, there is a δ such that $f(\mathcal{B}_r(p)) \subset \mathcal{B}_\epsilon(f(p)) \subset G$. This means that,

$$f^{-1}(f(\mathcal{B}_r(p))) \subset f^{-1}(G).$$

Hence,

$$\mathcal{B}_r(p) \subset f^{-1}(f(\mathcal{B}_r(p))) \subset f^{-1}(G).$$

Now, assume that If G is an open subset of Y then $f^{-1}(G)$ is open in X. Fix a $p \in X$ and $\epsilon > 0$. Then $f(p) \in Y$. Consider the ball $\mathcal{B}_\epsilon(f(p))$ which is an open set in Y. Thus, $f^{-1}(\mathcal{B}_\epsilon(f(p)))$ is open in X. But since $f(p) \in \mathcal{B}_\epsilon(f(p))$, p is in $f^{-1}(\mathcal{B}_\epsilon(f(p)))$ which is open in X. Thus there is $\delta > 0$, such that $\mathcal{B}_\delta(p) \subset f^{-1}(\mathcal{B}_\epsilon(f(p)))$. Hence,

$$f(\mathcal{B}_\delta(p)) \subset f\left(f^{-1}(\mathcal{B}_\epsilon(f(p)))\right) \subset \mathcal{B}_\epsilon(f(p)).$$

Hence, f is continuous. $\square$

The composition of continuous function is also continuous.

Proposition 1.3.4. *If $f : X \to Z$ and $g : Z \to W$ are continuous functions, then $g \circ f : X \to W$ is also continuous.*

Proof. This follows from the fact that $(g \circ f)^{-1}(G) = f^{-1}\left(\left(g^{-1}(G)\right)\right)$ for any $G \subset W$. $\square$

Example 1.3.1. *Let (X, d) is any metric space and let $x_0 \in X$. Let us define $f : X \to \mathbb{R}$ by $f(x) = d(x, x_0)$. Then f is continuous on X. This is because $|f(x) - f(y)| = |d(x, x_0) - d(y, x_0)| \leq d(x, y)$.*

Proposition 1.3.5. *If (X, d) is a metric space an $A \subset X$, then*

$$|\text{dist}(x, A) - \text{dist}(y, A)| \leq d(x, y),$$

for all $x, y \in X$. Thus, for any non-empty subset $A \subset X$, the function $f : X \to \mathbb{R}$ defined by $f(x) = \text{dist}(x, A)$ is continuous.

Proof. Let p be any element of A (since A is non-empty). Then,

$$\begin{aligned}\operatorname{dist}(x,A) &\leq d(x,p)\\ &\leq d(x,y)+d(y,p).\end{aligned}$$

Thus, $d(y,p) \geq \operatorname{dist}(x,A) - d(x,y)$ for any $p \in A$. Hence $\operatorname{dist}(y,A) \geq \operatorname{dist}(x,A) - d(x,y)$ i.e. $\operatorname{dist}(x,A) - \operatorname{dist}(y,A) \leq d(x,y)$. Starting the same argurment with y instead of x we get the result. ☐

Using the above Proposition, we can prove a famous result due to Urysohn.

Theorem 1.3.2. If A, B are two disjoint closed subsets of X, then there is a continuous function $f : X \to \mathbb{R}$ having the following properties:

(1) $0 \leq f(x) \leq 1$ for all $x \in X$.
(2) $f(x) = 0$ for all $x \in A$.
(3) $f(x) = 1$ for all $x \in B$.

Proof. Define $f : X \to \mathbb{R}$ by,

$$f(x) = \frac{\operatorname{dist}(x,A)}{\operatorname{dist}(x,A)+\operatorname{dist}(x,B)}.$$

Since A, B are disjoint *closed* sets, we cannot have $\frac{0}{0}$ and so the function is well defined. It is crucial that A, B be closed sets. If not, say $A = (0,1)$ and $B = (1,2)$, then A, B are disjoint. But if $x = 1$ is taken, then both $\operatorname{dist}(x,A) = \operatorname{dist}(x,B) = 0$.

Moreover, since $\operatorname{dist}(x,A), \operatorname{dist}(x,B)$ are continuous functions of x, f defined above is continuous. When $x \in A$, $\operatorname{dist}(x,B) \neq 0$ but $\operatorname{dist}(x,A) = 0$. When $x \in B$, $\operatorname{dist}(x,B) = 0$, but $\operatorname{dist}(x,A) \neq 0$. Thus, all properties are satisfied. ☐

Corollary 1.3.2.1. *If F is a closed subset of X and G is an open set containing F, then there is a continuous function $f : X \to \mathbb{R}$ such that $0 \leq f(x) \leq 1$ for all x in X, $f(x) = 1$ when $x \in F$ and $f(x) = 0$ when $x \notin G$.*

Proof. Using the Theorem above with $A = G^{\mathsf{c}}$ and $B = F$, we get the result. ☐

Definition 1.3.2 (Homeomorphism)**.** *If (X,d) and (Y,ρ) are metric spaces, then a map $f : X \to Y$ is called a homeomorphism if f is bijective and both f, f^{-1} are continuous. Two metric spaces are said to be homeomorphic if there is a homeomorphism from one onto the other.*

Remark 1.3.1. *Note that a bijection $f : X \to Y$ is a homeomorphism precisely when a sequence (x_n) in X converges to $x \in X$ if and only if $f(x_n) \to f(x)$. A homeomorphism identifies the two spaces. That is, via a* *homeomorphic map f, they have the same convergent sequences, open sets, closed sets and continuous functions. It is easy to see that homeomorphisms define equivalent relations.*

Another close concept is the notion of isometry.

Definition 1.3.3 (Isometry)**.** *Two metric spaces (X, d) and (Y, ρ) are isometric if there is a bijective map $f : X \to Y$ satisfying,*

$$\rho(f(a), f(b)) = d(a, b),$$

for every $a, b \in X$. Such a map f is called any isometric map.

Remark 1.3.2. *Note that if f is an isometric map, then f is injective. Let $f(a) = f(b)$, then $0 = \rho(f(a), f(b)) = d(a, b)$ means that $a = b$. If two metric spaces are isometric, then they are also homeomorphic. This is easy to see because an isometric map is continuous and (x_n) in X converges to $x \in X$ if and only if $f(x_n) \to f(x)$.*

Definition 1.3.4 (Equivalent metrics)**.** *If X is a set, then the two metrics d and ρ are said to be equivalent if they define the same convergent sequences. Equivalently, d, ρ are equivalent if the identity map $i : (X, d) \to (Y, \rho)$ is a homeomorphism.*

The following Proposition gives a useful way to get an equivalent metric.

Proposition 1.3.6. *For any metric space (X, d),*

$$\rho(x, y) := \frac{d(x, y)}{1 + d(x, y)},$$

defines an equivalent metric.

Proof. The proof is highly illustrative of an important technique. Note that positive reflexivity and symmetricity is easily verifiable. First let us investigate the function $f(t) = \frac{t}{1+t}$ for all $t > -1 \in \mathbb{R}$. Clearly, $f(t) < 1$ for all $t > -1$ and so f is bounded. Also $f'(t) = \frac{1}{1+t} - \frac{t}{(1+t)^2} = \frac{1}{(1+t)^2} > 0$. Hence f is increasing. $f''(t) = -2(1+t)^{-3} < 0$ and hence f' is decreasing. For any $s, t > 0$, let $g(t) = f(s)+f(t)-f(s+t)$. Then, $g'(t) = f'(t)-f'(s+t)$ and hence $g'(t) > 0$ since f' is decreasing. Thus g is an increasing function and thus for any $t > 0$, $g(t) > g(0) = 0$. Hence $f(s+t) \leq f(s) + f(t)$ for all $t \geq 0$.

Thus,

$$\begin{aligned}\rho(x,z) &= \frac{d(x,z)}{1+d(x,z)}\\ &\leq \frac{d(x,y)+d(y,z)}{1+d(x,y)+d(y,z)}\\ &= f\left[d(x,y)+d(y,z)\right]\\ &\leq f(d(x,y))+f(d(y,z))\\ &= \frac{d(x,y)}{1+d(x,y)}+\frac{d(y,z)}{1+d(y,z)}.\end{aligned}$$

For any sequence (x_n), such that $x_n \to x$ w.r.t d, we have $d(x_n,x) \to 0$ and so, $\rho(x_n,x) \to 0$. Also, $d(x,y) = \frac{1}{1-\rho(x,y)}$, which is defined since $\rho(x,y) < 1$, and hence if $x_n \to x$ w.r.t ρ, we have $x_n \to x$ w.r.t d too.

□

Remark 1.3.3. *Two equivalent metrics don't necessarily have the same Cauchy sequences.*

We need a stronger notion of continuity that preserves Cauchy sequences.

Definition 1.3.5 (uniform continuity)**.** *A function $f : (X,d) \to (Y,\rho)$ between two metric spaces is uniformly continuous if for any $\epsilon > 0$ there is a $\delta > 0$ such that for all x, y in X, if $d(x,y) < \delta$ then, $\rho(f(x), f(y)) < \epsilon$.*

Remark 1.3.4. *The crucial point in the above definition is that the choice of δ only depends on ϵ and not on any $x \in X$. Ofcourse, uniform continuity makes sense only for a set. There is no such thing as pointwise uniformly continuous.*

Theorem 1.3.3. The following concerns uniformly continuous function $f : (X,d) \to (Y,\rho)$ between two metric spaces.

(1) If (x_n) is a Cauchy sequence in X, the $(f(x_n))$ is a Cauchy sequence in Y.
(2) If $A \subset X$ and (Y,ρ) is complete, then f can be extend to a uniformly continuous function $f : \operatorname{cl} A \to Y$.

Proof. We prove in order.

(1) Fix any $\epsilon > 0$. We know that there is a $\delta > 0$ such that for any $x, y \in X$, $\rho(f(x), f(y)) < \epsilon$ whenever $d(x,y) < \delta$. Since, (x_n) is cauchy, there is an N such that for all $m, n \geq N$, $d(x_m, x_n) < \delta$. Hence, $\rho(f(x_m), f(x_n)) < \epsilon$ for all $m, n \geq N$. Thus, $(f(x_n))$ is Cauchy.

(2) First we need to extend f for all $x \in \operatorname{cl} A - A$. If $A = \operatorname{cl} A$, we are done. Hence, we can assume $A \neq \operatorname{cl} A$. Let $x \in \operatorname{cl} A$ such that $x \notin A$. There exists a sequence $(a_n) \in A$ such that $a_n \to x$. f map the sequence (a_n) to a sequence $(f(a_n)) \in Y$. Every convergent sequence is Cauchy and hence (a_n) is Cauchy and by the result above (since f is uniformly continuous), $(f(a_n))$ is Cauchy. Since, Y is complete, $(f(a_n)) \to y$, for some $y \in Y$. Let us define $f(x) = y$. Is f, well defined i.e. if (a_n), (b_n) are sequences in X such that $a_n \to x$ and $b_n \to x$, is it the case that $f(x) = y$ i.e. if $f(b_n) \to z$, we need to show $z = y$. Let $c_n = a_n$ if n is even an $c_n = b_n$ if n is odd. Thus $f(c_n)$ has one convergent subsequence that converges to y and another that converges to z. By uniqueness of limits, these must be equal.

We need to show that the extension of f is uniformly continuous on $\operatorname{cl} A$. To recap, our extension of f is given by $g\ : X \to Y$ such that $g\big|_A = f$ and for any $x \in \operatorname{cl} A$ such that $x \notin A$, we define $g(x) := \lim\limits_{n\to\infty} f(a_n)$ for any sequence (a_n) that converges to x. Let x, y be element of $\operatorname{cl} A$ that are not in A. Fix an $\epsilon > 0$. Then, we know that there is a sequence $a_n \to x$ and $b_n \to y$. Moreover, since f is uniformly continuous in A, there is a $\delta > 0$ such that $\rho(f(a_n), f(b_n)) < \epsilon$, whenever $d(a_n, b_n) < \delta$. Let $\delta' = \frac{\delta}{3}$.

$$\rho(g(x), g(y)) \leq \rho(g(x), f(a_n)) + \rho(f(a_n), f(b_n)) + \rho(f(b_n), g(y)).$$

Since, $a_n \to x$ there is a N_1 such that whenever $n \geq N_1$, $d(a_n, x) < \frac{\delta}{3}$. By definition, when $n \geq N_1$ we can make $\rho(f(a_n), g(x)) < \frac{\epsilon}{3}$. Similarly when $n \geq N_2$, we can make $\rho(f(b_n), g(y)) < \frac{\epsilon}{3}$. For $n \geq \max(N_1, N_2)$,

$$d(a_n, b_n) \leq d(a_n, x) + d(x, y) + d(y, b_n),$$

$d(a_n, b_n) < \delta$, whenever $d(x, y) < \delta'$. Since f is uniformly continuous, $\rho(f(a_n), f(b_n)) < \frac{\epsilon}{3}$ and hence $\rho(g(x), g(y)) < \epsilon$. Thus, we get the result.

□

1.4. Compactness

Definition 1.4.1. *Let X be a metric space. A family $\mathscr{G} := \{G_\alpha \subset X\ :\ \alpha \in A\}$ is said to* cover *a set S if S is contained in the union of all the $G_{\alpha' s}$. A family $\mathscr{U}$ is called a subcover if $\mathscr{U} \subset \mathscr{G}$ and $\mathscr{U}$ also covers S. A cover is called an open cover if it consists of open sets in X.*

Definition 1.4.2. *A set K in a metric space S is compact if every open cover of K contains a finite subcover.*

Proposition 1.4.1. *Let X be a metric space and let $K \subset X$ be compact. Then,*

(1) K is closed and bounded.

(2) For any closed subset $F \subset K$, F is compact.

(3) Every infinite subset of K has a limit point in K.

(4) If Y is any metric space, then for any continuous function $f : K \to Y$, $f(Y)$ is compact.

Proof. We prove in order.

(1) First we will show that K is bounded. Let x be any point in X. Then

$$K \subset X \subset \bigcup_{n=1}^{\infty} \mathcal{B}_n(x).$$

Since, K is compact, there are integers $i_1, i_2, \ldots, i_N$ such that

$$K \subset \bigcup_{k=1}^{N} \mathcal{B}_{i_k}(x).$$

Let $R = \max\{i_k : 1 \leq k \leq N\}$. Then, $K \subset \mathcal{B}_R(x)$; and hence, K is bounded.

To show K is closed we will show that its complement is open. Let $x \notin K$. For any $p \in K$, let us define $r_p = \frac{d(p,x)}{2}$. Let $G_p = \mathcal{B}_{r_p}(p)$. Then, $\mathscr{G} = \{G_p : p \in K\}$ is an open cover for K and so, there is finite subcover,

$$K \subset \bigcup_{i=1}^{n} \mathcal{B}_{r_{p_i}}(p_i).$$

Let $r = \min\{r_{p_i} : 1 \leq i \leq n\}$. Then, $r > 0$ and $\mathcal{B}_r(x)$ doesn't intersect K. Hence, K^{c} is open.

(2) See Rudin

(3) See Rudin

(4) See Rudin

□

An important corollary of (3) is that every sequence in K has a convergent subsequence. This property is called sequential compactness. Thus, we have shown that compactness implies sequential compactness.

Definition 1.4.3. *A subset K in a metric space is sequentially compact if for any sequence in K, there is a convergent subsequence in K.*

Definition 1.4.4. *We say that a subset K in a metric space is totally bounded if for any radius r there are finite points $x_1, x_2, \ldots, x_n$ in K such that*

$$K \subset \bigcup_{i=1}^{n} \mathcal{B}_r(x_i).$$

Part 2

Measure Theory

Chapter 2

The problem of measure: prescribing volume

One of the most fundamental concepts in geometry is that of the generalized *volume* or **measure** of a solid body E in one or more dimensions. In one dimension, this refers to *length* of the body E, and in two and three dimension we refer to the measure of E as the *area* and *volume* respectively. We can think of the measure as a **set function**, i.e. we input a solid body and it outputs its volume. A solid body can be abstracted as a set in $\mathbb{R}^n$. The Question is, can we prescribe a measure to *any* set in $\mathbb{R}^n$? Classically, to get the volume of any body E, one would find a sequence of approximations from within E and from outside and thus get some lower and upper bounds on the volume. In trying to generalize this to arbitrary sets in $\mathbb{R}^n$ we will follow the same procedure. Let us say the set function we are after is denoted by μ. In order for it to give us a reasonable volume we must expect μ to behave reasonably. By this, we mean

(1) If E is an interval $[a, b]$ in $\mathbb{R}$ or a rectangle $[a_1, b_1] \times [a_2, b_2]$, then $\mu(E)$ must equal its length $b - a$ in $\mathbb{R}$ or area $(b_1 - a_1) \times (b_2 - a_2)$ in $\mathbb{R}^2$.

(2) (Additivity) If E_1, E_2 are disjoint then $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$.

(3) (Translation invariance) If we translate E, we must not change its measure, that is $\mu(E + h) = \mu(E)$ for any h.

Figure 2.1. Rectangles in $\mathbb{R}^3, \mathbb{R}^2, \mathbb{R}$.

We will see that even such a basic requirement on our set function μ proves to be troublesome. In other words, we will show that there are sets in $\mathbb{R}^n$ that fail to satisfy the above requirements. We do not want to give up on any of these requirements. However, we will give up the requirement that all sets be measurable.

We start with measuring sets whose *volume* we already know from elementary geometry.

Definition 2.0.1 (Intervals and rectangles)**.** *An interval I is a subset of $\mathbb{R}$ of the form $[a,b] = \{x \in \mathbb{R} : a \le x \le b\}$, $[a,b) = \{x \in \mathbb{R} : a \le x < b\}$, $(a,b] = \{x \in \mathbb{R} : a < x \le b\}$ or $(a,b) = \{x \in \mathbb{R} : a < x < b\}$. We define the volume (length) of an interval I by $\lambda(I) = b - a$. We also denote this by $|I|$. A rectangle $R \in \mathbb{R}^n$ is a Cartesian product of n intervals, $R = \prod_{i=1}^{n} I_i$. We define the volume of a rectangle R as $\prod_{i=1}^{n} \lambda(I_i)$.*

For example a closed rectangle given by,

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n],$$

has a volume given by,

$$\lambda(R) = \lambda([a_1, b_1] \times \cdots \times [a_n, b_n]) = \prod_{i=1}^{n} (b_i - a_i).$$

Definition 2.0.2 (Elementary set)**.** *An elementary set is any subset of $\mathbb{R}^n$ which is the union of a finite number of disjoint rectangles.*

Note that any rectangle is an elementary set. It is easy to see that if E, F are elementary sets in $\mathbb{R}^n$, then $E \cup F$, $E \cap F$, $E - F$ are also elementary sets. How do we define the volume of an elementary set? Since an elementary set is a collection of finite disjoint rectangles, it makes sense that the volume of an elementary set is just the sum of the volumes of its component rectangles. Thus,

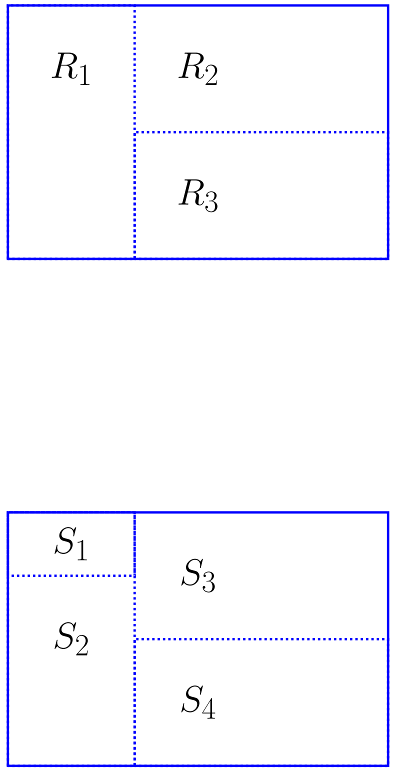

Figure 2.2. Measure of an elementary set. We show two decomposition of the elementary set.

Theorem 2.0.1 (Measure of an elementary set)**.** Let $E \subset \mathbb{R}^n$ be an elementary set. We define the (volume) measure of E as,

$$\mu(E) = \sum_{i=1}^{m} \lambda(R_i),$$

where $E = \dot{\bigcup}_{i=1}^{m} R_i$. Moreover, this measure is well defined, i.e. if $E = \dot{\bigcup}_{j=1}^{k} S_j$ its measure remains the same.

Proof. First we will show that for any interval I, we can compute its length by the following formula

$$\lambda(I) = \lim_{N\to\infty} \frac{1}{N} \#(I \cap \frac{1}{N}\mathbb{Z}),$$

Where $\frac{1}{N}\mathbb{Z} = \left\{\frac{n}{N} \;:\; n \in \mathbb{Z}\right\}$ and $\#$ refers to the cardinality of the finite set $I \cap \frac{1}{N}\mathbb{Z}$. It suffices to prove this for the unit interval $I = [0,1]$. For any N, $I \cap \frac{1}{N}\mathbb{Z}$ is the set $\{\frac{0}{N}, \frac{1}{N}, \ldots, \frac{N}{N}\}$ and thus $\#I \cap \frac{1}{N}\mathbb{Z} = N + 1$. Hence $\lim\limits_{N\to\infty} \frac{1}{N}\#(I \cap \frac{1}{N}\mathbb{Z})$ is equal to $\lim\limits_{N\to\infty} \frac{N+1}{N} = 1 = \lambda(I)$. See 2.3. Thus for any

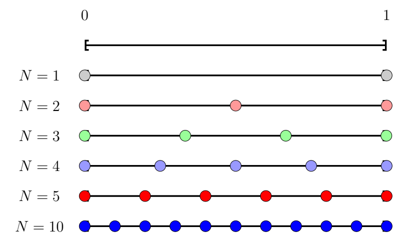

Figure 2.3. Discretization of the unit interval.

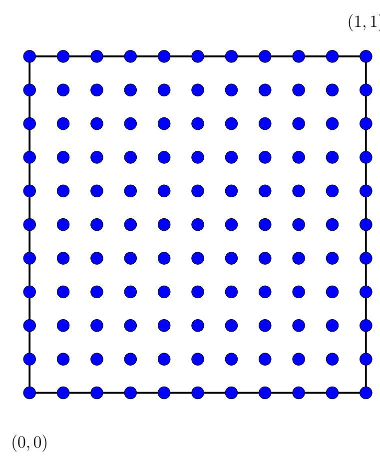

Figure 2.4. Discretization of the unit square.

rectangle $R_i \subset \mathbb{R}^n$, we have

$$\lambda(R_i) = \lim_{N\to\infty} \frac{1}{N^n} \#(R_i \cap \frac{1}{N}\mathbb{Z}^n).$$

Thus,

$$\lambda(E) = \sum_{i=1}^{m} \lambda(R_i) = \lim_{N\to\infty} \frac{1}{N^n} \#(E \cap \frac{1}{N}\mathbb{Z}^n),$$

which remains same no matter what is the decomposition of E. $\square$

We cannot extend this formula as the measure for any set in $\mathbb{R}^n$ since the limit may fail to exist. However, for elementary sets we can observe some important properties.

Property 1 If E, F are disjoint elementary sets then $\mu(E \cup F) = \mu(E) + \mu(F)$.

Property 2 If $E = \emptyset$, then $\mu(E) = 0$.

Property 3 If $E = R$, where R is a rectangle in $\mathbb{R}^n$, then $\mu(E) = \lambda(R)$.

Property 4 If E, F are elementary sets such that $E \subset F$ then $\mu(E) \leq \mu(F)$.

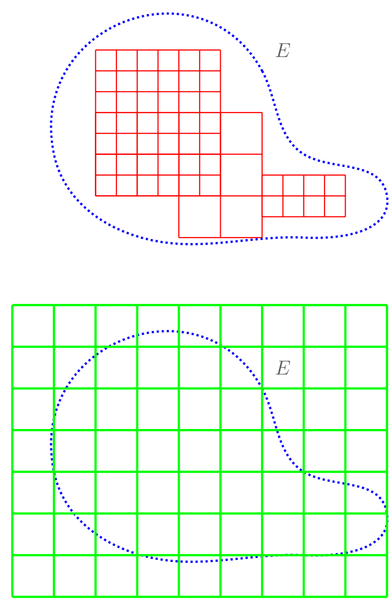

Figure 2.5. Approximating from within and without by an elementary set.

How do we extend this notion of measure to arbitrary sets in $\mathbb{R}^n$? We can follow the ideas first expounded by the Greeks i.e by approximating an arbitrary set by elementary sets from **inside** and **outside**. In other words we can define an approximate **outer** measure as follows:

$$\mu^*(E) = \inf_{B \supset E, B \text{ elementary}} \mu(B).$$

That is we *cover* E by an elementary set B and then take the infimum over all the measure of such B. Since $B = \bigcup_{i=1}^{n} R_i$, this is equivalent to saying,

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{n} \lambda(R_i) \ : \ \bigcup_{i=1}^{n} R_i \supset E\right\}.$$

Similary we approximate E from inside by

$$\mu_*(E) = \sup\left\{\sum_{j=1}^{n} \lambda(S_j) \ : \ \bigcup_{j=1}^{n} R_j \subset E\right\}.$$

Whenever these two coincide we say that $\mu(E) = \mu_*(E) = \mu^*(E)$. See 2.5. This is the construction by Jordan and appeals to the classical approximation ideas. However, there is a problem with such a construction. In analysis, sets occur as a limit of a sequence of sets and as such taking $\inf, \sup$ over finite unions of rectangles might prove to be trouble some. Hence, we tinker with this approximation and allow for a countable union of rectangles. To get a motiviation for using countable rectangles we show the decomposition theorem for an open set in $\mathbb{R}$.

Proposition 2.0.1 (Decomposition of an open interval)**.** *Any open subset $G \subset \mathbb{R}$ is a countable union of disjoint open intervals.*

Proof. Let G be an open set in $\mathbb{R}$. Then for any $x \in G$, there is an interval around x that is entirely contained in G. We wish to find the maximal interval containing x. To this end, let

$$a_x = \inf \{a < x \ : \ (a, x) \subset G\},$$
$$b_x = \sup \{x < b \ : \ (x, b) \subset G\},$$

with possibly infinite values for a_x, b_x. We take $I_x = (a_x, b_x)$. Clearly $x \in I_x$ and for any set J such that $x \in J$, we must have $J \subset I_x$. Then,

$$G = \bigcup_{x \in G} I_x.$$

Now suppose $I_x \cap I_y \neq \emptyset$. Thus there is an x such that $x \in I_x$ and $x \in I_y$ and thus $x \in I_x \cup I_y$. Since I_x is the maximal interval containing x we must have $I_x \cup I_y \subset I_x$. Similarly $I_x \cup I_y \subset I_y$. This is only possible when $I_x = I_y$. Thus we have a disjoint union. To show that the union is countable, we note that every interval I_x contains a rational number r_x. Since I_x, I_y are disjoint, $r_x \neq r_y$. Thus we can index each I_x with the corresponding rational number r_x. Note, there are an infinite number of rational numbers in each I_x. To pick one, we invoke the Axiom of Choice. Hence, we have the result. $\square$

It would make sense to define the measure of an open set $G \subset \mathbb{R}$ to be $\mu(G) = \sum_{i=1}^{\infty} \lambda(I_{x_i})$. Thus, we need a countable collection of rectangles instead of using just a finite collection. Unfortunately this decomposition doesn't quite extend to higher dimension. Fortunately, there is a similar decompostion for open sets in $\mathbb{R}^n$ that works for almost disjoint rectangles. Let us first define almost disjoint.

Definition 2.0.3 (Almost disjoint rectangles)**.** *We will denote the* interior *of a rectangle R by R°. Two rectangles R_i, R_j are almost disjoint if ${R_i}^\circ \cap {R_j}^\circ = \emptyset$.*

A cube $Q \subset \mathbb{R}^n$ is a **special** rectangle whose sides are all equal. It turns out that any open set in $\mathbb{R}^n$ can be written as a countable collection of *almost* disjoint cubes. To prove this, we will define *dyadic* cubes in $\mathbb{R}^n$.

Note that for any $x \in \mathbb{R}$ and for any $M \in \mathbb{Z}^+$, by the archimedes principle there is an integer k such that, $\frac{k}{M} \leq x < \frac{(k+1)}{M}$ i.e k is the greatest integer of Mx. If we choose $M = 2^N$ for any $N \in \mathbb{Z}^+$ we get a dyadic partition of $\mathbb{R}$ for every N. We can extend this idea to $\mathbb{R}^n$.

Definition 2.0.4 (Dyadic cube). *A (closed) dyadic cube $C_{\mathbf{k},N} \in \mathbb{R}^n$ is given by*

$$C_{\mathbf{k},N} = \left\{ \boldsymbol{x} \in \mathbb{R} \ : \ \frac{k_i}{2^N} \leq x_i \leq \frac{k_i + 1}{2^N} \quad 1 \leq i \leq n \right\},$$

where $\boldsymbol{k} \in \mathbb{Z}^n$.

Let $\mathscr{C}_n$ be the collection of all dyadic cubes. Note that $\mathscr{C}_n \subset \mathscr{J}_n$, where $\mathscr{J}_n$ is the family of all closed rectangles. For a fixed N, each side of the cube is of length $\frac{1}{2^N}$. Hence the volume of the cube is $\left(\frac{1}{2^N}\right)^n$. Note that for any $\boldsymbol{x}, \boldsymbol{y} \in C_{\mathbf{k},N}$ the distance $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \frac{\sqrt{n}}{2^N}$. For clarity, we will denote $\mathscr{C}_n^N$ to be the collection of all dyadic cubes of level N in $\mathbb{R}^n$. Two dyadic cubes are almost disjoint if they intersect only in the boundary.

Proposition 2.0.2. *Every open set in $\mathbb{R}^n$ is a countable union of almost disjoint cubles.*

Proof. Let G be an open set. First we will show that there is a countable collection of almost disjoint dyadic cubes $\mathscr{Q} = \{Q_i \in \mathscr{C}_n \ : \ i \in \mathbb{Z}^+\}$ such that,

$$\bigcup_{i=1}^{\infty} Q_i \subset G.$$

We start with a collection of dyadic cubes of level 0. We pick $Q_i \in \mathscr{C}_n^0$ such that $Q_i \subset G$. Next, we identify Q_k such that $Q_k \cap G \neq \emptyset$. Then we find $Q_j \subset Q_k$ such that $Q_j \in \mathscr{C}_n^1$. If $Q_j \subset G$ we add it to the collection $\mathscr{Q}$ otherwise we repeat by finding $Q_j \in \mathscr{C}_n^2$ and so on. Since by construction each $Q_i \subset G$ we get $\bigcup\limits_{i=1}^{\infty} Q_i \subset G$. This is illustrated in 2.6.

Now consider an $\boldsymbol{x} \in G$. Since G is open there is an $\epsilon > 0$ such that the open ball $\mathcal{B}_\epsilon(\boldsymbol{x}) \subset G$. Now, there is a $C_{\mathbf{k},N}$ such that $\boldsymbol{x} \in C_{\mathbf{k},N}$, however $C_{\mathbf{k},N}$ may or may not be completely in G. But we can successively increase i.e., we find an $M > N$ such that $C_{\mathbf{k},M} \subset C_{\mathbf{k},N}$ and $C_{\mathbf{k},M} \subset G$. Thus it must be in $\mathscr{Q}$ and hence,

$$G \subset \bigcup_{i=1}^{\infty} Q_i.$$

This is illustrated in 2.7. $\square$

If we want to include countable collection of rectangles, we want to make sure that we are not changing the elementary measure of a rectangle i.e if a rectangle is a countable union of almost disjoint rectangles then its volume measure must equal the infinite sum of its component rectangles. We collect

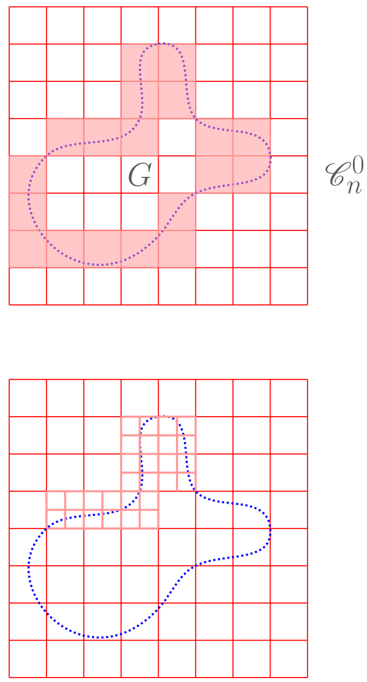

Figure 2.6. Illustration of *paving* of $G \subset \mathbb{R}^n$ by dyadic cubes. The cubes Q_k filled in red are such that $Q_k \cap G \neq \emptyset$. Thus we increase the level.

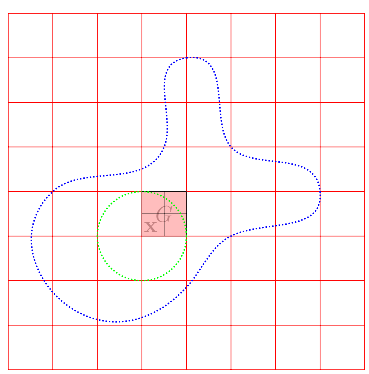

Figure 2.7. Illustration of finding a dyadic cube containing $\boldsymbol{x}$ and inside $\mathcal{B}_\epsilon(\boldsymbol{x})$.

all these facts in the following proposition. Note that since a cube is also a rectangle, the following results also hold for a cube. To make the ideas clear we will show these facts for $\mathbb{R}^2$. The proposition is valid for all $\mathbb{R}^d, d \geq 1$.

Proposition 2.0.3. *Let R be a two dimensional rectangle with sides parallel to the co-ordinate axes. Then,*

(1) *Suppose that* $R = I_1 \times I_2$ *where each* $I_i \subset [a, b] \subset \mathbb{R}$. *If each* I_i *is an almost disjoint union of closed, bounded intervals i.e*

$$\{I_{i,j_i} \subset [a,b] \ : \ 1 \leq j_i \leq N_i\}, \quad \textit{and}$$

$$I_i = \bigcup_{j_i=1}^{N_i} I_{i,j_i}.$$

Define the rectangles,

$$S_{j_1,j_2} = I_{1,j_1} \times I_{2,j_2},$$

then,

$$\lambda(R) = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \lambda(S_{j_1,j_2}).$$

(2) *If a rectangle* R *is an almost disjoint, finite union of rectangles,* $\{R_i \ : \ 1 \leq i \leq N\}$, *then*

$$\lambda(R) = \sum_{i=1}^{N} \lambda(R_i).$$

Note that this is another proof of 2.0.1.

(3) *If a rectangle* R *is a subset of another rectangle* S, *then*

$$\lambda(R) \leq \lambda(S).$$

(4) *If a rectangle* R *is* covered *by finite union of rectangles,* $\{R_i \ : \ 1 \leq i \leq N\}$, *then*

$$\lambda(R) \leq \sum_{i=1}^{N} \lambda(R_i).$$

(5) *If a rectangle* R *is the countable union of almost disjoint rectangles,* $\{R_i \ : \ i \in \mathbb{Z}^+\}$, *then*

$$\lambda(R) = \sum_{i=1}^{\infty} \lambda(R_i).$$

Proof. We prove in order,

(1) Let us denote the length of an interval I by $|I|$, then since I is an elementary set we can see that, (for example using the formula in 2.0.1)

$$|I_i| = \sum_{j_i=1}^{N_i} |I_{i,j_i}|.$$

By definition,

$$\lambda(R) = |I_1||I_2|.$$

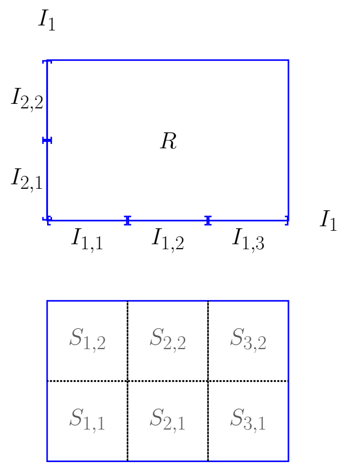

Figure 2.8. Illustration of proof 2.0.3 (1).

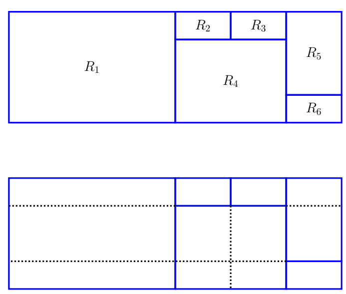

Figure 2.9. Illustration of proof 2.0.3 (2).

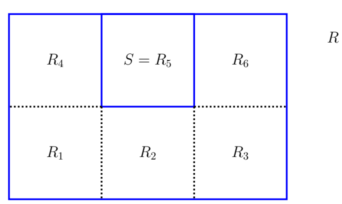

Figure 2.10. Illustration of proof 2.0.3 (3).

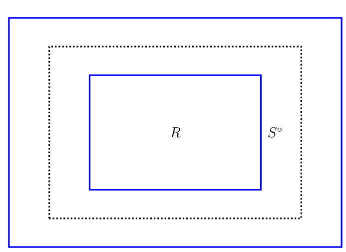

Figure 2.11. Illustration of proof 2.0.3 (5).

Thus,

$$\begin{aligned}\lambda(R) &= |I_1||I_2| \\ &= \sum_{j_1=1}^{N_1} |I_{1,j_1}| \sum_{j_2=1}^{N_2} |I_{2,j_2}| \\ &= \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} |I_{1,j_1}||I_{2,j_2}| \\ &= \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \lambda(S_{j_1,j_2})\end{aligned}$$

See 2.8.

(2) If R is the almost disjoint union of R_i for $1 \leq i \leq N$, then we can extend the sides of each R_i. Each $\lambda(R_i)$ can then be expressed as a sum in (1). Also sides of R have been partitioned and its volume measure can be expressed as a sum in (1). Easy to see that $\lambda(R) = \sum_{i=1}^{N} \lambda(R_i)$. See 2.9.

(3) Note that both R, S are elementary sets and we have see that volume measure preserves monotonicity. Here we will give a different proof. If $R \subset S$, then extend the sides of R to intersect the sides of S. Then using the result above $\lambda(S)$ is the sum of almost disjoint rectangles $\{R_i\}$, where one of R_i is equal to R. See 2.10. Hence we get the result.

(4) If $R \subset \bigcup_{i=1}^{N} R_i$, then we get, $R = \bigcup_{i=1}^{N} (R \cap R_i)$. However, each $R \cap R_i$ may not be disjoint. Let us denote by $S_1 = R \cap R_1$ and subsequent $S_j = (R \cap R_j) - (S_1 \cup \ldots S_{j-1})$. Then R is the almost disjoint finite union of rectangles S_j and by the result in (2) $\lambda(R) = \sum_{j=1}^{N} \lambda(S_j)$. But each $S_j \subset R_j$ and hence by (3) above $\lambda(S_j) \leq \lambda(R_j)$. Thus we get the result.

(5) Note that R *covers* any finite almost disjoint union $\bigcup_{i=1}^{N} R_i$ i.e, $\bigcup_{i=1}^{N} R_i \subset R$. Thus from above (roles reversed),

$$\lambda\left(\bigcup_{i=1}^{N} R_i\right) \leq \lambda(R).$$

But since we have an almost finite disjoint union, using (2),

$$\sum_{i=1}^{N} \lambda(R_i) \leq \lambda(R).$$

Taking the limit, we get

$$\lambda(R) \geq \sum_{i=1}^{\infty} \lambda(R_i).$$

To prove the other inequality we have to use a compactness argument.

Let us extend each R_i to get an open covering of R. That is we get an S_i such that,

$$\lambda(S_i) \leq \lambda(R_i) + \frac{\epsilon}{2^i}.$$

and ${S_i}^\circ \supset R_i$. (Note this means that we cover R_i by S_i which itself can be covered by something larger than R_i since each R_i is bounded.) See 2.11. Thus we have an open cover of R given by, $\mathscr{G} = \{{S_i}^\circ : i \in \mathbb{Z}^+\}$. Since we are in a compact space there is a finite subcover $\mathscr{G}_N$ such that, $R \subset \bigcup_{i=1}^{N} S_i$. Hence, $\lambda(R) \leq \sum_{i=1}^{N} \lambda(S_i)$. Taking the limit to ∞ we get the result.

□

Thus we have seen, that to measure open sets we need a countable covering of elementary sets. But not all sets are open sets. Since we are describing an abstract set function, we must be willing to accept strange sets. The most famous strange set is the **Cantor** set.

To construct a Cantor set we begin with the unit interval $[0, 1]$ and let C_1 denote the set obtained from deleting the middle third open interval from $[0, 1]$, that is

$$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1].$$

To get to the next level, we delelte the middle third open interval from each of the above intervals to get,

$$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1].$$

This procedure yields a sequence of closed sets (C_k), where each C_k is a finite union of closed interval of length $\frac{1}{3^k}$, and $C_1 \supset C_2 \supset C_3 \ldots$.

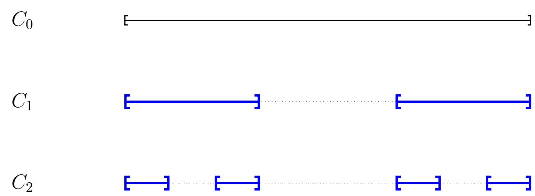

Figure 2.12. First three levels of Cantor set in $\mathbb{R}$.

Definition 2.0.5. *The Cantor set C is a compact set in $\mathbb{R}$ defined by,*

$$C = \bigcap_{k=1}^{\infty} C_k,$$

where C_k is the set obtained by removing $2^k - 1$ middle third open intervals as described above.

Proposition 2.0.4. *Let C be the Cantor set in $\mathbb{R}$. Then,*

(1) C does not contain any open intervals.

(2) Every point of C is a limit point of C.

(3) C is uncountable.

Proof. We prove (1), (2).

(1) Let us assume that there is an interval (a, b) such that $(a, b) \subset C$. Then for any $n \in \mathbb{Z}^+$, $(a, b) \subset C_n$. But since C_n contains 2^n disjoint closed intervals of length $1/3^n$, (a, b) must be a subset of one of these. But if we take n large enough such that $1/3^n < (b - a)$ we get a contradiction.

(2) Fix an $\epsilon > 0$. Since $x \in C$, $x \in C_n$ for every $n \in \mathbb{Z}^+$. But C_n consists of disjoint closed intervals of length $1/3^n$ and hence x belongs to one of these closed interval. If x is the left end of the interval pick y to be the right end of that interval, if x is the right end of the interval pick y to be the left end of that interval and in any other case pick y to be the left end of the interval. Thus for any n we have found a $y \in C$ such that $|y - x| < 1/3^n$. Pick n large enough such that $1/3^n < \epsilon$, then $|y - x| < \epsilon$. Since ϵ was arbitrary and $y \neq x$, we have shown that x is a limit point of C and thus there exists a sequence in C that converges to x. $\square$

What is the measure of the Cantor set? Approximating C from **inside** seems tricky since C doesn't contain any open intervals. In what follows, we will approximate from outside by allowing a countable collection of rectangles. As a convenience, we take the rectangles to be cubes. Nothing is lost in such a construction.

Uptill now, we have a notion of measure for elementary sets in $\mathbb{R}^d$. We now construct a set function that approximates *any* set from **outside**. Such a construction will yield the familiar measure when the set is an elementary set. Let us give a precise definition,

Definition 2.0.6 (Outer (volume) measure). *If E is any subset of $\mathbb{R}^d$, the outer measure of E is,*

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \lambda(Q_i) \ : \ E \subset \bigcup_{i=1}^{\infty} Q_i\right\}.$$

Hence, $\mu^* : \mathscr{P}(\mathbb{R}^d) \to [0, \infty]$.

Example 2.0.1. *What is the outer measure of a point $\boldsymbol{x} \in \mathbb{R}^d$? A point $\boldsymbol{x}$ is covered by $Q = \prod_{i=1}^{d} I_i$, where $I_i = [x_i, x_i]$, and its volume is equal to 0.*

Example 2.0.2. *What is the outer measure of a cube Q? Since $Q \subset Q$, from the definition of outer measure $\mu^*(Q) \leq \lambda(Q)$. Now fix an ϵ, by the definition of infimum and outer measure, there is a countable collection of cubes $\{Q_i\}$, such that $Q \subset \bigcup_{i=1}^{\infty} Q_i$ and*

$$\sum_{i=1}^{\infty} \lambda(Q_i) < \mu^*(Q) + \epsilon.$$

By extending the sides of Q_j we can find a S_j such that $S_j{}^\circ \supset Q_j$ and $\lambda(S_j) \leq \lambda(Q_j) + \frac{\epsilon}{2^j}$. See 2.11. Hence $Q \subset \bigcup_{j=1}^{\infty} S_j{}^\circ$ and since Q is compact (closed and bounded), there exists a finite subcover such that (after renumbering)

$$Q \subset \bigcup_{j=1}^{N} S_j{}^\circ \subset \bigcup_{j=1}^{N} S_j.$$

From 2.0.3 (3), we have

$$\lambda(Q) \leq \sum_{j=1}^{N} \lambda(S_j).$$

Thus,

$$\lambda(Q) \leq \sum_{j=1}^{N} \lambda(Q_j) + \frac{\epsilon}{2^j} \leq \sum_{j=1}^{\infty} \lambda(Q_j).$$

Hence $\lambda(Q) \leq \mu^*(Q)$. Therefore we get that $\lambda(Q) = \mu^*(Q)$. In fact, we can show that for any rectangle R its volume measure equals the outer measure.

Now we are ready to list some of the properties of outer measure.

Proposition 2.0.5 (Properties of (volume) outer measure)**.** *Let μ^* be the outer measure defined for any subset in $\mathbb{R}^d$. Then,*

Property 1 *(Monotonicity) If $E_1 \subset E_2$, then*

$$\mu^*(E_1) \leq \mu^*(E_2).$$

Property 2 *(Countable sub-additivity) If $E = \bigcup\limits_{j=1}^{\infty} E_j$, then*

$$\mu^*(E) \leq \sum_{j=1}^{\infty} \mu^*(E_j).$$

Property 3 *(Approximation by open sets) If $E \subset \mathbb{R}^d$, then*

$$\mu^*(E) = \inf\{\mu^*(G) \ : \ E \subset G,\ G \text{ open}\}.$$

Property 4 *If $E = E_1 \cup E_2$ and $d(E_1, E_2) > 0$, then*

$$\mu^*(E) = \mu^*(E_1) + \mu^*(E_2).$$

Property 5 *If E is the countable union of almost disjoint cubes, i.e $E = \bigcup\limits_{i=1}^{\infty} Q_i$, then*

$$\mu^*(E) = \sum_{j=1}^{\infty} \lambda(Q_j).$$

Proof. We prove in order.

(1) The collection of cubes that cover E_2 also cover E_1. Thus if $\mathscr{Y}_1$ is the collection of cubes covering E_1 and $\mathscr{Y}_2$ is the collection of cubes covering E_2, then $\mathscr{Y}_1 \supset \mathscr{Y}_2$. Hence taking infimum of the sums of the volumes in the respective families, we get the result.

(2) Fix an ϵ. Then for each j there is a collection of cubes $\{Q_i^j\}$ such that $E_j \subset \bigcup\limits_{i=1}^{\infty} Q_i^j$ and,

$$\sum_{i=1}^{\infty} \lambda\left(Q_i^j\right) < \mu^*(E_j) + \frac{\epsilon}{2^j}.$$

We can observe that,

$$E \subset \bigcup_{i,j} Q_i^j,$$

and hence,

$$\begin{aligned}
\mu^*(E) &\leq \sum_i \sum_j \lambda\left(Q_i^j\right) \\
&= \sum_j (\sum_i \lambda\left(Q_i^j\right)) \\
&< \sum_j (\mu^*(E_j) + \frac{\epsilon}{2^j}) \\
&\leq \sum_{j=1}^{\infty} \mu^*(E_j) + \epsilon
\end{aligned}$$

Since, ϵ was arbitrary we get the result.

(3) If there is a G such that $G \supset E$, then by mononoticity $\mu^*(G) \geq \mu^*(E)$ and thus $\mu^*(E)$ is less than or equal to the infimum of the outer measure of all such open sets G such that $G \supset E$. For the other direction, fix an $\epsilon > 0$. Hence, there is a collection of cubes $\{Q_{\epsilon,i}\}$ such that $E \subset \bigcup_{i=1}^{\infty} Q_\epsilon, i$ and,

$$\sum_{i=1}^{\infty} \lambda(Q_{\epsilon,i}) < \mu^*(E) + \epsilon.$$

We can extend each $Q_{\epsilon,i}$ to get $\tilde{Q}_{\epsilon,i}$ such that,

$$\lambda\left(\tilde{Q}_{\epsilon,i}^{\circ}\right) < \lambda(Q_{\epsilon,i}) + \frac{\epsilon}{2^i}.$$

Let $G_\epsilon = \bigcup_{i=1}^{\infty} \tilde{Q}_{\epsilon,i}^{\circ}$. Then G_ϵ is open and from monotonicity,

$$\mu^*(G_\epsilon) \leq \sum_{i=1}^{\infty} \mu^*\left(\tilde{Q}_{\epsilon,i}^{\circ}\right) = \sum_{i=1}^{\infty} \lambda\left(\tilde{Q}_{\epsilon,i}^{\circ}\right).$$

Thus,

$$\begin{aligned}
\mu^*(G_\epsilon) &\leq \sum_{i=1}^{\infty} \lambda\left(\tilde{Q}_{\epsilon,i}^{\circ}\right) \\
&\leq \sum_{i=1}^{\infty} (\lambda(Q_{\epsilon,i}) + \frac{\epsilon}{2^i}) \\
&< \mu^*(E) + 2\epsilon,
\end{aligned}$$

since, ϵ was arbitrary we get the result.

(4) If $E = E_1 \cup E_2$, then from monotonicity $\mu^*(E) \leq \mu^*(E_1) + \mu^*(E_2)$. Thus we need to show the other inequality.

Fix an $\epsilon > 0$. There is a collection of cubes $\mathscr{Q} = \{Q_i\}$ such that $E \subset \bigcup_{i=1}^{\infty} Q_i$ and

$$\sum_{i=1}^{\infty} \lambda(Q_i) < \mu^*(E) + \epsilon.$$

Let us construct a countable collection that covers E. Consider the set E_1. If a cube Q_i intersects E_1 only and not E_2 we add i to an index set J_1. If a cube Q_i intersects both E_1 and E_2, then we refine the cube until it intersects only E_1. This can be done since if a cube intersects both E_1 and E_2 its diameter must be larger than $d(E_1, E_2)$. Thus refining the cube (by a higher dyadic level) we can get it to intersect only one. Thus we have an index set J_1 such that $E_1 \subset \bigcup_{i \in J_1} Q_i$, where each $Q_i, i \in J_1$ is a subset of cubes in $\mathscr{Q}$. Similarly we can find a J_2 such that $E_2 \subset \bigcup_{i \in J_2} Q_i$. Thus,

$$\begin{aligned}
\mu^*(E_1) + \mu^*(E_2) &\leq \sum_{i \in J_1} \mu^*(Q_i) + \sum_{i \in J_2} \mu^*(Q_i) \\
&= \sum_{i \in J_1} \lambda(Q_i) + \sum_{i \in J_2} \lambda(Q_i) \\
&\leq \sum_{Q_i \in \mathscr{Q}} \lambda(Q_i) \\
&< \mu^*(E) + \epsilon.
\end{aligned}$$

(5) From countable sub-additivity, $\mu^*(E) \leq \sum_{i=1}^{\infty} \lambda(Q_i)$. For the other direction, we can **reduce** the sides of each cube Q_i to get a cube $\tilde{Q}_i$ such that $\tilde{Q}_i$ is strictly a subset of Q_i and,

$$\lambda(Q_i) \leq \lambda\left(\tilde{Q}_i\right) + \frac{\epsilon}{2^i}.$$

For any an finite number N then we have $d(\tilde{Q}_i, \tilde{Q}_j) > 0$ where i, j range over N. Hence, we can use the result above to deduce that for any N,

$$\mu^*\left(\bigcup_{i=1}^{N} \tilde{Q}_i\right) = \sum_{i=1}^{N} \lambda\left(\tilde{Q}_i\right) \geq \sum_{i=1}^{N} (\lambda(Q_i) - \frac{\epsilon}{2^j}).$$

Note that $\epsilon(\frac{1}{2} + \cdots + \frac{1}{2^N}) < \epsilon$ for any N and $\bigcup_{i=1}^{N} \tilde{Q}_i \subset E$. Thus,

$$\mu^*(E) \geq \sum_{i=1}^{N} \lambda(Q_i) - \epsilon.$$

Since this is true for any N, we can take the limit $N \to \infty$ to get,

$$\sum_{i=1}^{\infty} \lambda\left(Q_i\right) \leq \mu^*\left(E\right) + \epsilon.$$

Thus we get the result.

□

It is crucial to observe that we do not have additivity for arbitrary disjoint sets. That is if E_1, E_2 are disjoint then we cannot conclude that $\mu^*\left(E_1 \cup E_2\right) = \mu^*\left(E_1\right) + \mu^*\left(E_2\right)$. For example take $E_k = \emptyset$ for $k > N$ in 2.0.5 (2) above and note that $\mu^*\left(\emptyset\right) = 0$. This is only true when E_1 and E_2 are **metrically** disjoint. Also from the last property, using 2.0.2 the outer measure of an open set written as the union of a countable collection of almost disjoint cubes is the infinite sum of the volume of those cubes.

Since additivity is a crucial property we need identify those sets that preserve additivity, more importantly countable additivity. The notion of measurability isolates a collection of subsets in $\mathbb{R}^d$ for which the exterior measure statisfies all our desired properties including countable additivity. Since open sets satisfy countable additivity, we make the following criteria for measurablity.

Definition 2.0.7 ((Volume) Lebesgue measure in $\mathbb{R}^n$)**.** *A subset $E \subset \mathbb{R}^d$ is Lebesgue measurable or simply measurable, if for any $\epsilon > 0$ there is an open set G such that $E \subset G$ and,*

$$\mu^*\left(G - E\right) \leq \epsilon.$$

When this is true we denote the measure of E by $\mu\left(E\right) = \mu^*\left(E\right)$.

We will show that the collection of measurable sets behave nicely w.r.t. set operations like unions, intersection, complements. In particular, we will show that a countable union of measurable sets are measurable, countable intersection of measurable sets are measurable and the complement of measurable set is measurable. Moreover, the measure of elementary set is the usual volume and we preserve (countable) additivity of disjoint sets and translation invariance. However we will also show that the class of measurable sets is strictly smaller than $\mathscr{P}\left(\mathbb{R}^d\right)$.

Before we list the properties of measurable sets, we will list a useful lemma.

Lemma 2.0.1. *If $F \subset \mathbb{R}^d$ is closed and $K \subset \mathbb{R}^d$ is compact and these two sets are disjoint, then $d(F, K) > 0$.*

Proof. Let $x \in K$. Then $x \notin F$ and hence there is a $\delta_x > 0$ such that $\mathcal{B}_{\delta_x}\left(x\right) \cap F = \emptyset$. Hence for any $y \in F$, $d(x, y) > \delta_x$. By choosing a smaller

δ_x we can make $d(x,F) \geq 2\delta_x$. Now the collection $\{\mathcal{B}_{\delta_x}(x) : x \in K\}$ is an open cover for K and since K is compact there is a finite subcover $\left\{\mathcal{B}_{\delta_{x_i}}(x_i) : 1 \leq i \leq N\right\}$. Let $\delta = \min\{\delta_{x_i} : 1 \leq i \leq N\}$. Then for any $x \in K$, there is a x_j such that $d(x, x_j) < \delta_j$. Pick any $y \in F$, then $d(y,x) \geq d(y,x_j) - d(x,x_j) > \delta$ and hence $d(F,K) > 0$. $\square$

Theorem 2.0.2 (Properties of Lebesgue (volume) measure)**.** Let μ be the Lebesgue measure and let $\mathscr{L}$ be the collection of all (Lebesgue) measurable sets. We observe the following:

Property 1 If G is an open subset of $\mathbb{R}^d$, then $G \in \mathscr{L}$.

Property 2 (Completeness) If $\mu^*(E) = 0$, then $E \in \mathscr{L}$. In particular if F is a subset of a set of exterior measure 0, then $F \in \mathscr{L}$.

Property 3 If $\{E_i\}$ is a countable collection of measurable sets, then

$$\bigcup_{i=1}^{\infty} E_i \in \mathscr{L}.$$

Property 4 If F is a closed subset of $\mathbb{R}^d$, then $F \in \mathscr{L}$.

Property 5 If $E \in \mathscr{L}$, then $E^{\mathsf{c}} \in \mathscr{L}$.

Property 6 If $\{E_i\}$ is a countable collection of measurable sets, then

$$\bigcap_{i=1}^{\infty} E_i \in \mathscr{L}.$$

Property 7 If $\{E_i\}$ is a countable collection of pairwise disjoint measurable sets, then

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

Note that a set E has Lebesgue measure zero if $\mu(E) = 0$. We have a useful characterization of sets of Lebesgue measure zero.

Proposition 2.0.6 (Lebesgue measure 0)**.** *A set $E \subset \mathbb{R}^n$ has Lebesgue measure zero iff for any $\epsilon > 0$ there is a sequence of rectangles $\{R_i\} \subset \mathbb{R}^n$ with their sides parallel to the co-odrdinate axes such that,*

$$E \subset \bigcup_{i=1}^{\infty} R_i \quad \textit{and} \quad \sum_{i=1}^{\infty} \lambda(R_i) < \epsilon$$

Chapter 3

Elements of measure theory: measurable sets

In this chapter, we will define the abstract notion of measurability of sets.

3.1. Sigma Algebra of sets

Definition 3.1.1 (Power set)**.** *Let X be a non-empty set. The family of all subsets of X is called the* power set *of X and is denoted by $\mathscr{P}(X)$.*

Definition 3.1.2 (Increasing sequence of sets)**.** *Let (A_n) be a sequence of subsets of X. If $A_1 \subset A_2 \subset A_3 \ldots$ and $\bigcup\limits_{i \in \mathbb{Z}^+} A_i = A$, then we say that the A_n form an increasing sequence of sets and increase to A. We denote this by $A \nearrow A$.*

Definition 3.1.3 (Decreasing sequence of sets)**.** *Let (A_n) be a sequence of subsets of X. If $A_1 \supset A_2 \supset A_3 \ldots$ and $\bigcap\limits_{i \in \mathbb{Z}^+} A_i = A$, then we say that the A_n form an decreasing sequence of sets and decrease to A. We denote this by $A_n \searrow A$.*

Given two subsets $A, B \subset X$, we can write the union as a disjoint union as follows:

$$A \bigcup B = (A) \dot{\bigcup} (B \cap A^{\mathsf{c}})$$

Here we use $\dot{\bigcup}$ to denote that the union is between elements that are disjoint.

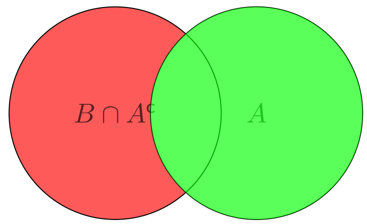

Figure 3.1. Disjoint union of two sets

The above observation along with the *DeMorgan's* Law can be used to show the following statements that equate an arbitrary union to disjoint union.

Remark 3.1.1. *Let (A_n) be a sequence of subsets of X. Then,*

(1) $\bigcup\limits_{i=1}^{n} A_i = (A_1) \dot{\bigcup} (A_2 \cap A_1^{\mathsf{c}}) \dots \dot{\bigcup} \left(A_n \cap A_{n-1}^{\mathsf{c}} \cdots \cap A_1^{\mathsf{c}}\right)$. *Let us construct new sets $F_k = (A_k) \bigcap \left(\bigcup_{i=1}^{k-1} A_i\right)^{\mathsf{c}}$ for $k > 1$ and $F_1 = A_1$. Thus $\bigcup\limits_{i \in \mathbb{Z}^+} A_i = \dot{\bigcup}_{i \in \mathbb{Z}^+} F_i$. Note that F_k is just $A_k - \bigcup_{i=1}^{k-1} A_i$. In this way, we can get a sequence of pairwise disjoint sets (F_n) whose countable union is the same as the countable union of the originial sequence.*

(2) *If $A_n \nearrow A$ then,*
   (a) $A_n = \bigcup\limits_{i=1}^{n} A_i = A_1 \dot{\bigcup} (A_2 - A_1) \dots \dot{\bigcup} (A_n - A_{n-1})$,
   (b) $A = \bigcup\limits_{i \in \mathbb{Z}^+} A_i = \dot{\bigcup}_{i \in \mathbb{Z}^+} (A_i - A_{i-1})$, *where we take $A_0 = \emptyset$.*

(3) *If $A_n \nearrow A$ then $A_n^{\mathsf{c}} \searrow A^{\mathsf{c}}$. If $A_n \searrow A$ then $A_n^{\mathsf{c}} \nearrow A^{\mathsf{c}}$.*

(4) *We can construct a sequence (B_n) such that $B_n \nearrow A$ where $A = \bigcup\limits_{n=1}^{\infty} A_n$ as follows. $B_1 = A_1$, $B_2 = A_1 \cup A_2$ and so on, where $B_n = \bigcup\limits_{k=1}^{n} A_k$. This is an important trick to generate an increasing sequence of sets whose countable unions are the same.*

Definition 3.1.4 (Algebra of sets)**.** *Let X be a non-empty set. An* algebra *on X is a* non-empty *collection of sets $\mathscr{A} \subset \mathscr{P}(X)$ with the following properties,*

(1) If $E \in \mathscr{A}$ then $E^{\mathsf{c}} \in \mathscr{A}$.
(2) If $E_1, E_2, \ldots, E_n \in \mathscr{A}$ then $\bigcup_{i=1}^{n} E_i \in \mathscr{A}$.

Definition 3.1.5 (Sigma-Algebra of sets)**.** *Let X be a non-empty set. A σ- algebra on X is a* non-empty *collection of sets $\mathscr{A} \subset \mathscr{P}(X)$ with the following properties,*

(1) If $E \in \mathscr{A}$ then $E^{\mathsf{c}} \in \mathscr{A}$.
(2) If $(E_n) \in \mathscr{A}$ then $\bigcup\limits_{i \in \mathbb{Z}^+} E_i \in \mathscr{A}$.

Theorem 3.1.1 (Properties of sigma algebras)**.** If $\mathscr{A}$ is a $\sigma -$ algebra on X then,

Property 1 If $E_1, E_2, E_3, \ldots, E_n \in \mathscr{A}$ then $\bigcup\limits_{i=1}^{n} E_i \in \mathscr{A}$.
Property 2 $X, \emptyset \in \mathscr{A}$.
Property 3 If $(E_n) \in \mathscr{A}$ then $\bigcap\limits_{i \in \mathbb{Z}^+} E_i \in \mathscr{A}$.
Property 4 If $E_1, E_2, E_3, \ldots, E_n \in \mathscr{A}$ then $\bigcap\limits_{i=1}^{n} E_i \in \mathscr{A}$.
Property 5 If $A, B \in \mathscr{A}$ then $A - B \in \mathscr{A}$.

Note that for an algebra the closure under countable intersection is generally not true. Thus a $\sigma -$ algebra on a set X is an algebra that is closed under countable union and intersection.

Proof. We prove in order,

(1) Let $E_i = E_n$ for $i > n$. Then $\bigcup\limits_{i=1}^{n} E_i = \bigcup\limits_{i \in \mathbb{Z}^+} E_i \in \mathscr{A}$.

(2) Since $\mathscr{A}$ is non empty there is a $E \subset X$ such that $E \in \mathscr{A}$. Thus $E^{\mathsf{c}} \in \mathscr{A}$. Thus $X = E \cup E^{\mathsf{c}} \in \mathscr{A}$ from the above property. Since $X \in \mathscr{A}, X^{\mathsf{c}} = \emptyset \in \mathscr{A}$.

(3) From DeMorgan's Law.

(4) Since $E_1, E_2, \ldots, E_n \in \mathscr{A}, E_1^{\mathsf{c}}, \ldots, E_n^{\mathsf{c}} \in \mathscr{A}$. From (1), $\bigcup\limits_{i=1}^{n} E_i^{\mathsf{c}} \in \mathscr{A}$.

Thus its complement $\bigcap\limits_{i=1}^{n} E_i$ is in $\mathscr{A}$.

(5) From above, $A - B = A \cap B^{\mathsf{c}}$ which is in $\mathscr{A}$.

□

Example 3.1.1. *The following are all σ- algebras.*

(1) $\mathscr{P}(X)$ *is a* σ*- algebra. It is called the maximal* σ*- algebra on* X*.*

(2) $\{\emptyset, X\}$ *is called the minimal* σ*- algebra on* X*.*

(3) For any $B \subset X$*, the collection* $\{\emptyset, B, B^{\mathsf{c}}, X\}$ *is a* σ*- algebra.*

(4) The collection $\mathscr{A} = \{A \subset X \ : \ A \text{ is countable or } A^{\mathsf{c}} \text{ is countable}\}$ *is a* σ*- algebra.*

Proof. $(A^{\mathsf{c}})^{\mathsf{c}} = A$. Thus if $A \in \mathscr{A}$, it is either countable or A^{c} is countable. Thus $A^{\mathsf{c}} \in \mathscr{A}$. Let $(A_n) \in \mathscr{A}$. If all are countable then the countable union is countable and so is in $\mathscr{A}$. If not there exist an index $j \in \mathbb{Z}^+$ such that A_j^{c} is countable. Thus $(\bigcup\limits_{i \in \mathbb{Z}^+} A_i)^{\mathsf{c}} = \bigcap\limits_{i \in \mathbb{Z}^+} A_i^{\mathsf{c}} \subset A_j^{\mathsf{c}} \in \mathscr{A}$. □

*(5) (**Restricted*** σ*- algebra) Let* $E \subset X$ *be any set and let* $\mathscr{A}$ *be a* σ*- algebra on* X*. Then the collection* $\mathscr{A}_E = \{E \cap A \ : \ A \in \mathscr{A}\}$ *is* σ*- algebra on* E*.*

Proof. Let $E_1 \in \mathscr{A}_E$, then $E_1 = E \cap A$ for some $A \in \mathscr{A}$. The complement of E_1 in E is given by $E \cap E_1^{\mathsf{c}}$. Thus we have $E \cap E_1^{\mathsf{c}} = E \cap (E^{\mathsf{c}} \cup A^{\mathsf{c}})$ and so $E \cap E_1^{\mathsf{c}} = (E \cap E^{\mathsf{c}}) \cup (E \cap A^{\mathsf{c}})$. Thus $E \cap E_1^{\mathsf{c}} = E \cap A^{\mathsf{c}} \in \mathscr{A}_E$. Let (E_n) be a sequence in $\mathscr{A}_E$. Each $E_i = E \cap A_i$ for some $A_i \in \mathscr{A}$. Therefore $\bigcup\limits_{i \in \mathbb{Z}^+} E_i = E \cap \bigcup\limits_{i \in \mathbb{Z}^+} A_i \in \mathscr{A}_E$. □

Theorem 3.1.2 (Equivalent Characterization of σ- algebra)**.** An algebra of sets $\mathscr{A}$ is a σ- algebra on X iff it is closed under complements and for any sequence of pairwise disjoint sets contained in the collection their union is also contained in the collection.

Proof. $\Rightarrow$ is immediately evident from the definition of σ- algebra.

$\Leftarrow$ Consider a sequence of sets $(E_n) \in \mathscr{A}$. Using 3.1.1, we can construct a sequence of pairwise disjoint sets (F_n), such that $\bigcup\limits_{i \in \mathbb{Z}^+} A_i = \dot{\bigcup}_{i \in \mathbb{Z}^+} F_i \in \mathscr{A}$. We need to $\mathscr{A}$ to be an algebra so that $F_n = A_n \cap A_1^{\mathsf{c}} \cap \cdots \cap A_{n-1}^{\mathsf{c}}$ is in $\mathscr{A}$. □

Definition 3.1.6 (Generated σ- algebra)**.** *Let* X *be any non-empty set and let* $\mathscr{E} \subset \mathscr{P}(X)$*. The* σ*- algebra generated by* $\mathscr{E}$ *is the unique smallest* σ*- algebra containing* $\mathscr{E}$ *and is denoted by* $\sigma(\mathscr{E})$*.*

In set theory a *smallest* set in a collection is the set that is contained in every other set of the collection. The next theorem guarantees the existence of a smallest σ- algebra.

Theorem 3.1.3 (Sigma Algebra generated by an arbitrary collection). Let X be a non-empty set,

(1) The intersection of any collection of σ- algebra on X is itself a σ- algebra.

(2) Let $\mathscr{E} \subset \mathscr{P}(X)$. There is a unique smallest σ- algebra $\sigma(\mathscr{E})$ containing $\mathscr{E}$ in the sense that any σ- algebra containing $\mathscr{E}$ must contain $\sigma(\mathscr{E})$.

Proof. Given a non-empty set X,

(1) Let $\mathcal{G}$ be a collection of σ- algebra on X. Then,

$$\mathscr{F} = \bigcap_{\mathscr{A} \in \mathcal{G}} \mathscr{A} = \{E \subset X \ : \ \forall (\mathscr{A} \in \mathcal{G}), E \in \mathscr{A}\}.$$

Let $E \in \mathscr{F}$. Thus $\forall (\mathscr{A} \in \mathcal{G})$, $E \in \mathscr{A}$ and hence $E^{\mathsf{c}} \in \mathscr{A}$ for all $\mathscr{A} \in \mathcal{G}$. Thus $E \in \mathscr{F}$. Similary for a sequence of sets (E_n) in $\mathscr{F}$, we have that $\bigcup_{i \in \mathbb{Z}^+} E_i \in \mathscr{F}$.

(2) Let $\mathcal{G} = \{\mathscr{A} \text{ on } X \ : \ \mathscr{E} \subset \mathscr{A}\}$. Then $\mathcal{G}$ is a collection of σ- algebra on X. Define $\sigma(\mathscr{E}) = \bigcap_{\mathscr{A} \in \mathcal{G}} \mathscr{A}$. From above, $\sigma(\mathscr{E})$ is a σ- algebra. By definition for any σ- algebra $\mathscr{A}$ on X that contains $\mathscr{E}$, $\mathscr{A} \in \mathcal{G}$ and so $\sigma(\mathscr{E}) \subset \mathscr{A}$.

$\square$

Remark 3.1.2. *We make a few observations about the generated σ- algebra.*

(1) If $\mathscr{E}$ is a σ- algebra, then $\mathscr{E} = \sigma(E)$.

Proof. Note that $\mathscr{E} \subset \mathscr{E}$ and so if $\mathscr{E}$ is a σ- algebra, from the proof above $\sigma(\mathscr{E}) \subset \mathscr{E}$. That $\mathscr{E} \subset \sigma(\mathscr{E})$ is evident from the definition of $\sigma(\mathscr{E})$. $\square$

(2) For any $A \subset X$, we have $\sigma(\{A\}) = \{\emptyset, A, A^{\mathsf{c}}, X\}$.

(3) If $\mathscr{E} \subset \mathscr{F}$, then $\sigma(\mathscr{E}) \subset \sigma(\mathscr{F})$.

Proof. $\mathscr{E} \subset \mathscr{F} \subset \sigma(\mathscr{F})$. Thus $\sigma(\mathscr{E}) \subset \sigma(\mathscr{F})$. $\square$

A very important principle to prove statements about generated sigma algebra is called the *principle of good sets* which is described in the following remark:

Remark 3.1.3 (Principle of good sets). *Suppose we want to show that for any collection $\mathscr{E}$, the generated sigma algebra $\sigma(\mathscr{E})$ has a certain property P. IF we can find a sub-collection $\mathscr{F} \subset \sigma(\mathscr{E})$ that has this property P, and show that $\mathscr{F}$ is a σ- algebra, then $\sigma(\mathscr{E})$ has property P, provided $\mathscr{E} \subset \mathscr{F}$.*

We can generate new σ- algebra from existing σ- algebra using inverse maps.

Theorem 3.1.4 (Pre-Image σ- algebra)**.** Let $f : X \to Y$ be a function.

(1) If $\mathscr{A}$ is a σ- algebra on Y, then
$$f^{-1}(\mathscr{A}) := \left\{f^{-1}(A) \ : \ A \in \mathscr{A}\right\}$$
is a σ- algebra on X.

(2) If $\mathscr{B}$ is a σ- algebra on X, then
$$\left\{A \subset Y \ : \ f^{-1}(A) \in \mathscr{B}\right\}$$
is a σ- algebra on Y.

(3) If $\mathscr{E}$ is a collection of sets in Y, then
$$\sigma\left(f^{-1}(\mathscr{E})\right) = f^{-1}(\sigma(\mathscr{E}))$$

Proof. Given a function $f : X \to Y$,

(1) Let $E \in f^{-1}(\mathscr{A})$. Thus $E = f^{-1}(A)$ for some $A \in \mathscr{A}$. $E^{\mathsf{c}} = \left(f^{-1}(A)\right)^{\mathsf{c}} = f^{-1}(A^{\mathsf{c}})$ and thus $E^{\mathsf{c}} \in f^{-1}(\mathscr{A})$. Let (E_n) be a sequence of sets in $f^{-1}(\mathscr{A})$. Then $E_i = f^{-1}(A_i)$. Thus $\bigcup\limits_{i\in\mathbb{Z}^+} E_i = \bigcup_{i\in\mathbb{Z}^+} f^{-1}(A_i) = f^{-1}\left(\bigcup\limits_{i\in\mathbb{Z}^+} A_i\right) \in f^{-1}(\mathscr{A})$.

(2) Let E be in the collecion, hence, $E \subset Y$ such that $f^{-1}(E) \in \mathscr{B}$, hence $f^{-1}(E^{\mathsf{c}}) = \left(f^{-1}(E)\right)^{\mathsf{c}} \in \mathscr{B}$. Thus E^{c} is in the collection. Similar argument for countable union.

(3) We need to show $\sigma\left(f^{-1}(\mathscr{E})\right) \subset f^{-1}(\sigma(\mathscr{E}))$ and $f^{-1}(\sigma(\mathscr{E})) \subset \sigma\left(f^{-1}(\mathscr{E})\right)$.

For any $A \in \mathscr{E}$, $A \in \sigma(\mathscr{E})$ and hence $f^{-1}(\mathscr{E}) \subset f^{-1}(\sigma(\mathscr{E}))$. But from (1), we know that $f^{-1}(\sigma(\mathscr{E}))$ is a σ- algebra, and thus $\sigma\left(f^{-1}(\mathscr{E})\right) \subset f^{-1}(\sigma(\mathscr{E}))$.

The other direction is tricky i.e.
$$f^{-1}(\sigma(\mathscr{E})) \subset \sigma\left(f^{-1}(\mathscr{E})\right).$$
We will use 3.1.3; that is, we will construct a sub-collection of $f^{-1}(\sigma(\mathscr{E}))$ which is a subset of $\sigma\left(f^{-1}(\mathscr{E})\right)$. What is a sub-collection of $f^{-1}(\sigma(\mathscr{E}))$ that satisfies this property? It is precisely a collection of some sets A in $\sigma(\mathscr{E})$ for which $f^{-1}(A)$ is in $\sigma\left(f^{-1}(\mathscr{E})\right)$. Thus consider the collection of good sets $\mathscr{F}$, such that
$$\mathscr{F} = \left\{A \in \sigma(\mathscr{E}) \ : \ f^{-1}(A) \in \sigma\left(f^{-1}(\mathscr{E})\right)\right\}$$
What we want to show is that for any $A \in \sigma(\mathscr{E})$ we should have,
$$f^{-1}(A) \in \sigma\left(f^{-1}(\mathscr{E})\right).$$
Clearly $\mathscr{F}$ is not empty because $\mathscr{E} \subset \mathscr{F}$. This is easily observed because $f^{-1}(\mathscr{E}) \subset \sigma\left(f^{-1}(\mathscr{E})\right)$. If we show that F is a σ- algebra

we are done. Let $E \in \mathscr{F}$. Therefore $f^{-1}(E) \in \sigma\left(f^{-1}(\mathscr{E})\right)$ but this means that $E^{\mathsf{c}} \in \mathscr{F}$ since $f^{-1}(E^{\mathsf{c}}) = \left(f^{-1}(E)\right)^{\mathsf{c}} \in \sigma\left(f^{-1}(\mathscr{E})\right)$. Similary for countable unions. Thus $\mathscr{F}$ is a σ- algebra containing $\mathscr{E}$ and so $\sigma(\mathscr{E}) \subset \mathscr{F}$. By construction, $\mathscr{F} \subset \sigma(\mathscr{E})$ and so $\sigma(\mathscr{E}) = \mathscr{F}$ and hence we are done.

□

Of particular interest is the situation involving a metric space, for example $\mathbb{R}^n$. In this case, we have a built in natural concept of distance plus the associated topology of sets, for example open sets, closed sets etc.

Definition 3.1.7 (Borel Sets)**.** *Let (X, d) be a metric space. The σ- algebra generated by the collection of open sets in X is called the* Borel σ- *algebra on X and is denoted by $\mathfrak{B}_X$. Its members are called the* Borel sets*.*

We can equivalently generate the Borel σ- algebra using the family of closed sets. The Borel sets include open sets, countable union an intersection of open sets, closed sets, countable union and intersection of closed sets and so on.

Definition 3.1.8. *A countable intersection of open sets is called a G_δ set, a countable union of closed sets is called a F_σ set, a countable union of G_δ set is called $G_{\delta\sigma}$ and so on.*

The Borel σ- algebra on $\mathbb{R}^n$, $\mathfrak{B}_{\mathbb{R}^n}$ is particularly important. When $n = 1$, we get the Borel σ- algebra on the real line, $\mathfrak{B}_{\mathbb{R}}$. The most interesting generators are the family of open and half-open rectangles,

$$
\begin{aligned}
\mathscr{I}_n &= \left\{[a_1, b_1] \times \dots [a_n, b_n] \;:\; a_j, b_j \in \mathbb{R}\right\}, \\
\mathscr{I}_n^o &= \left\{(a_1, b_1) \times \dots (a_n, b_n) \;:\; a_j, b_j \in \mathbb{R}\right\}, \\
\mathscr{I}_n^{cr} &= \left\{(a_1, b_1] \times \dots (a_n, b_n] \;:\; a_j, b_j \in \mathbb{R}\right\}, \\
\mathscr{I}_n^{cl} &= \left\{[a_1, b_1) \times \dots [a_n, b_n) \;:\; a_j, b_j \in \mathbb{R}\right\}. \\
\mathscr{I}_{\mathbb{Q}_n}^o &= \left\{(a_1, b_1) \times \dots (a_n, b_n) \;:\; a_j, b_j \in \mathbb{Q}\right\}.
\end{aligned}
$$

Let us denote by $\mathscr{G}, \mathscr{F}, \mathscr{K}$, the collection of open, closed and compact sets in $\mathbb{R}^n$. The following theorem is very useful in realizing the generating sets of $\mathfrak{B}_{\mathbb{R}^n}$.

Theorem 3.1.5 (Generating Borel Sets in $\mathbb{R}^n$)**.** We have,

(1) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{G})$.
(2) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{F})$.
(3) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{K})$.
(4) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{I}_n^o)$.
(5) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{I}_n^{cr})$.
(6) $\mathfrak{B}_{\mathbb{R}^n} = \sigma\left(\mathscr{I}_n^{cl}\right)$.
(7) $\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{I}_n)$.

Proof. (1) is just the definition of Borel sets. For any open set $U \in \mathscr{G}$, $U^{\mathsf{c}} \in \mathscr{F}$. Thus $U \in \sigma(F)$ i.e $\mathscr{G} \subset \sigma(\mathscr{F})$ and thus $\sigma(\mathscr{G}) \subset \sigma(\mathscr{F})$. A similar argument with roles reversed leads to $\sigma(\mathscr{F}) \subset \sigma(\mathscr{G})$. Any compact set K is also a closed set so $\sigma(\mathscr{K}) \subset \sigma(\mathscr{F})$. Let us write a closed set as a countable union of compact sets in $\mathbb{R}^n$. If $F \in \mathscr{F}$, then let $F_i = F \cap \overline{\mathcal{B}_i(\mathbf{0})}, i \in \mathbb{Z}^+$. Each F_i is an intersection of closed sets and is bounded and hence is compact in $\mathbb{R}^n$. Moreover,

$$F = \bigcup_{i \in \mathbb{Z}^+} F_i.$$

Thus $F \in \sigma(\mathscr{K})$ and so $\sigma(\mathscr{F}) \subset \sigma(\mathscr{K})$. Thus we have shown the following,

$$\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{G}) = \sigma(\mathscr{K}) = \sigma(\mathscr{F}).$$

Let $R \in \mathscr{I}_n^o$ be a open rectangle (box) in $\mathbb{R}^n$. Any open rectangle is an open set, thus $\sigma\left(\mathscr{I}_{\mathbb{Q}_n}^o\right) \subset \sigma(\mathscr{I}_n^o) \subset \sigma(\mathscr{G})$. Now for the other direction consider any open set $U \in \mathscr{G}$. $U = \bigcup_{p \in U} \mathcal{B}_{\epsilon_p}(p)$. But for any open ball we can inscribe a rectangle with rational endpoints in it and so,

$$U = \bigcup_{R \in \mathscr{I}_{\mathbb{Q}_n}^o; R \subset U} R$$

Thus $\mathscr{G} \subset \sigma\left(\mathscr{I}_{\mathbb{Q}_n}^o\right)$ and so we get the following,

$$\mathfrak{B}_{\mathbb{R}^n} = \sigma(\mathscr{I}_n^o)$$

Let $R^{cr} \in \mathscr{I}_n^{cr}$ be a half-open rectangle in $\mathbb{R}^n$ closed at right. Thus, $R^{cr} = (a_1, b_1] \times \ldots (a_n, b_n]$ which can be written as,

$$(a_1, b_1] \times \ldots (a_n, b_n] = \bigcap_{j \in \mathbb{Z}^+} (a_1, b_1 + 1/j) \times \ldots (a_n, b_1 + 1/j).$$

Thus, $\mathscr{I}_n^{cr} \subset \sigma(\mathscr{I}_n^o)$. Let $R \in \mathscr{I}_n^o$ be a open rectangle (box) in $\mathbb{R}^n$. Thus, $R = (a_1, b_1) \times \ldots (a_n, b_n)$ which can be written as,

$$(a_1, b_1) \times \ldots (a_n, b_n) = \bigcup_{j \in \mathbb{Z}^+} (a_1, b_1 - 1/j] \times \ldots (a_n, b_1 - 1/j].$$

Thus, $\mathscr{I}_n^o \subset \sigma(\mathscr{I}_n^{cr})$. Hence we have shown,

$$\sigma(\mathscr{I}_n^o) = \sigma(\mathscr{I}_n^{cr})$$

A similar argument can be done for the family $\mathscr{I}_n^{cl}$. □

Corollary 3.1.5.1. $\mathfrak{B}_{\mathbb{R}}$ *is generated by each of the following:*

(1) (Open intervals): $\mathscr{E}_1 = \{(a,b) \ : \ a < b\}$.
(2) (closed intervals): $\mathscr{E}_2 = \{[a,b] \ : \ a < b\}$.
(3) (half open closed right intervals): $\mathscr{E}_3 = \{(a,b] \ : \ a < b\}$
(4) (half open closed left intervals): $\mathscr{E}_4 = \{[a,b) \ : \ a < b\}$
(5) (open rays): $\mathscr{E}_5 = \{(a,\infty) \ : \ a \in \mathbb{R}\}$
(6) (open rays): $\mathscr{E}_6 = \{(-\infty,a) \ : \ a \in \mathbb{R}\}$
(7) (closed rays): $\mathscr{E}_7 = \{[a,\infty) \ : \ a \in \mathbb{R}\}$
(8) (closed rays): $\mathscr{E}_8 = \{(a,\infty] \ : \ a \in \mathbb{R}\}$

Before moving onto the subject of measures, we'll give a technical result that will be useful in constructing measure for the Borel sets. The following just abstracts the class of half open rectangles.

Definition 3.1.9 (Elementary family)**.** *Let X be a non-empty set. An elementary family is a collection $\mathscr{E}$ of subsets of X such that,*

(1) $\emptyset \in \mathscr{E}$.
(2) If $E, F \in \mathscr{E}$ *then* $E \cap F \in \mathscr{E}$.
(3) If $E \in \mathscr{E}$, *then* E^{c} *is a finite disjoint union of members of* $\mathscr{E}$.

Example 3.1.2. *Consider the family $\mathscr{I}_1^{cr}$ of half-open closed right interval in $\mathbb{R}$. If we include $-\infty, \infty$ then we refer to sets of this family as **h-intervals**. These are collection of subsets of $\mathbb{R}$ of the form $(a,b], (a,\infty), \emptyset$. $\mathscr{I}_1^{cr}$ is an elementary family. To check the last condition of the definition, consider the set $(a,b]$. Its complement is set $(-\infty,a] \cup (b,\infty)$ which is the disjoint union of two **h-intervals**.*

Theorem 3.1.6 (Constructing an algebra from elementary family)**.** If $\mathscr{E}$ is an elementary family, the collection, $\mathscr{A}$ of finite disjoint unions of members of $\mathscr{E}$ is an algebra.

Proof. First note that $\mathscr{A}$ is not empty since $\mathscr{E} \subset \mathscr{A}$. This is because any $E \in \mathscr{E}$ can be written as the disjoint union of $E \bigcup \emptyset$.

Let $A \in \mathscr{A}$. Then $A = E_1 \dot{\bigcup} E_2 \dot{\bigcup} \dots \dot{\bigcup} E_n$. We need to show that $A^{\mathsf{c}} \in \mathscr{A}$. Note that $A^{\mathsf{c}} = E_1^{\mathsf{c}} \cap E_2^{\mathsf{c}} \cap \cdots \cap E_n^{\mathsf{c}}$. However each $E_m^{\mathsf{c}} = E_1^m \dot{\bigcup} \dots \dot{\bigcup} E_{n_m}^m$.

Therefore A^{c} is,

$$\bigcap_{m=1}^{n} \dot{\bigcup}_{j=1}^{n_m} E_j^m = \dot{\bigcup} \left\{ E_{j_1}^1 \cap \cdots \cap E_{j_n}^n \ : \ 1 \le j_m \le n_m,\, 1 \le m \le n \right\},$$

which is a disjoint union of finite sets in $\mathscr{E}$ and hence in $\mathscr{A}$.

Let us show that $\mathscr{A}$ is closed under union of two sets. By induction we will get closure of finite union. Let $A, B \in \mathscr{A}$. Then $A = \dot{\bigcup}_{i=1}^{n_A} C_i$ and $B = \dot{\bigcup}_{i=1}^{n_B} D_i$. We need to show that $A \cup B \in \mathscr{A}$ i.e the following set,

$$A \cup B = \dot{\bigcup} \{ C_{j_A} \cup D_{j_B} \ : \ 1 \le j_A \le n_A; 1 \le j_B \le n_B \}.$$

Each $C_{j_A} \cup D_{j_B}$ belong to $\mathscr{A}$. To see this, let us show that if $C, D \in \mathscr{E}$ then $C \cup D \in \mathscr{A}$. Note that, $D^{\mathsf{c}} = \dot{\bigcup}_{i=1}^{J} E_i$. Thus $C \cap D^{\mathsf{c}} = \dot{\bigcup}_{i=1}^{J} C \cap E_i$. Since $C, E_i \in \mathscr{E}$ we get $C \cap D^{\mathsf{c}} \in \mathscr{A}$. But $C \cup D = (C \cap D^{\mathsf{c}}) \dot{\bigcup} (D)$. Since $(C \cap D^{\mathsf{c}})$ is the disjoint union of sets in $\mathscr{E}$, $C \cup D$ is then a disjoint union of sets in $\mathscr{E}$ and hence $C \cup D \in \mathscr{A}$. Thus from induction if $A_1, A_2, \ldots, A_n \in \mathscr{E}$ then $\bigcup\limits_{i=1}^{n} A_i \in \mathscr{A}$. Hence, each $C_{j_A} \cup D_{j_B}$ belong to $\mathscr{A}$. Thus by induction on $n_A \times n_B$ we get the result. Now we can induct on a finite sequence of sets $A_i \in \mathscr{A}$ to show that $\mathscr{A}$ is closed under finite unions. □

Remark 3.1.4. *Note that if $\mathscr{E}$ is the collection $\mathscr{J}_1^{cr}$, then the collection $\mathscr{A}$ of finite disjoint unions of **h-intervals** is an algebra by 3.1.6. But note that $\mathscr{E} \subset \mathscr{A}$ and thus $\sigma(\mathscr{E}) \subset \sigma(\mathscr{A})$. However, $\sigma(\mathscr{A}) = \mathfrak{B}_{\mathbb{R}}$ which is also equal to $\sigma(\mathscr{E})$ from 3.1.5.*

We close this section by looking at a few more important family of sets.

Definition 3.1.10 (Monotone Class)**.** *Let X be a non-empty set. A monotone class $\mathscr{E}$ is a non-empty collection of subsets of X which is closed under monotone sequence of sets, i.e.*

1. *If (A_n) is a sequence of subsets of X in $\mathscr{E}$ such that $A_i \subset A_{i+1}$ for each i, then $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{E}$.*
2. *If (A_n) is a sequence of subsets of X in $\mathscr{E}$ such that $A_i \supset A_{i+1}$ for each i, then $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{E}$.*

Proposition 3.1.1. *An algebra $\mathscr{A}$ is a sigma algebra if and only if it is a monotone class.*

Proof. Let $\mathscr{A}$ be a sigma algebra. Then since it is closed under countable unions, $\mathscr{A}$ is also a monotone class.

Let $\mathscr{A}$ be a monotone class and consider a sequence (A_n). From remark 3.1.1, we can construct an increasing sequence (B_n) such that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Since $\mathscr{A}$ is a monotone class, we get $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$. Hence $\mathscr{A}$ is a sigma-algebra. $\square$

Given a non-empty collection of sets $\mathscr{E}$, the minimal sigma algebra containing $\mathscr{E}$, is definied to be the sigma algebra generated by $\mathscr{E}$. Similarly, we can talk about the minimal monotone class containing $\mathscr{E}$ as the intersection of all monotone classes containing $\mathscr{E}$. We denote this by $\mathcal{M}(\mathscr{E})$. The next theorem is one of a type called monotone class theorems and its technique is quite illustrative of a useful proof technique in measure theory using remark 3.1.3.

Theorem 3.1.7. Let $\mathscr{E}$ be an algebra. Then, $\mathcal{M}(\mathscr{E}) = \sigma(\mathscr{E})$.

Proof. We need to show $\sigma(\mathscr{E}) \subset \mathcal{M}(\mathscr{E})$ and $\mathcal{M}(\mathscr{E}) \subset \sigma(\mathscr{E})$.

Since $\mathscr{E}$ is a field, $\sigma(\mathscr{E})$ is a monotone class containing $\mathscr{E}$, we get from minimality, $\mathcal{M}(\mathscr{E}) \subset \sigma(\mathscr{E})$.

For the other direction, if we show that $\mathcal{M}(\mathscr{E})$ is a sigma algebra containg $\mathscr{E}$, we are done, but this amounts to showing that $\mathcal{M}(\mathscr{E})$ is an algebra. Thus, we need to show: $\mathcal{M}(\mathscr{E})$ is an algebra, that is for any $A \in \mathcal{M}(\mathscr{E})$, $A^{\mathsf{c}} \in \mathcal{M}(\mathscr{E})$ and for any $A, B \in \mathcal{M}(\mathscr{E})$, $A \cup B \in \mathcal{M}(\mathscr{E})$. We will use the good sets principle. Let us define the following property that we want $\mathcal{M}(\mathscr{E})$ to satisfy.

$$P_1(A) := A^{\mathsf{c}} \in \mathcal{M}(\mathscr{E})\text{and } \forall\,(B \in \mathcal{M}(\mathscr{E}))\; A \cup B \in \mathcal{M}(\mathscr{E}).$$

Let,

$$\mathscr{F}_1 = \{A \in \mathcal{M}(\mathscr{E}) \;:\; P_1(A)\}\,.$$

By definition, $\mathscr{F}_1$ is an algebra and $\mathscr{F}_1 \subset \mathcal{M}(\mathscr{E})$. Thus, we need to show that $\mathscr{F}_1 \supset \mathcal{M}(\mathscr{E})$, which is equivalent to showing that

(1) $\mathscr{F}_1 \supset \mathscr{E}$ and,

(2) $\mathscr{F}_1$ is a monotone class.

To show (2), let (A_n) be an increasing sequence of sets in $\mathscr{F}_1$ and so $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}(\mathscr{E})$. Let B be any set in $\mathcal{M}(\mathscr{E})$, then $A_i \cup B$ is in $\mathscr{F}_1$. Hence, $(A_i \cup B)$ is an increasing sequence in $\mathcal{M}(\mathscr{E})$ and so

$$\left(\bigcup_{i=1}^{\infty} A_i\right) \cup B = \bigcup_{i=1}^{\infty} (A_i \cup B) \in \mathcal{M}(\mathscr{E}).$$

Thus, $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}_1$. Also $A_1^{\mathsf{c}} \supset A_2^{\mathsf{c}} \ldots$, and so,

$$(\bigcup_{n=1}^{\infty} A_n)^{\mathsf{c}} = \bigcap_{n=1}^{\infty} A_n^{\mathsf{c}} \in \mathcal{M}(\mathscr{E}).$$

Thus, (2) is satisfied.

Showing (1) is tricky and hence we will show,

$$\mathscr{F}_1 \supset \mathscr{F}_2 \supset \mathscr{E},$$

for a suitable $\mathscr{F}_2$ such that $\mathscr{F}_2$ is monotone. This would mean that $\mathscr{F}_1 \supset \mathscr{F}_2 \supset \mathcal{M}(\mathscr{E})$. To construct $\mathscr{F}_2$, we will define a new property:

$$P_2(A) := A^{\mathsf{c}} \in \mathcal{M}(\mathscr{E}) \text{and } \forall\, (B \in \mathscr{E})\; A \cup B \in \mathcal{M}(\mathscr{E}).$$

Let,

$$\mathscr{F}_2 = \{A \in \mathcal{M}(\mathscr{E}) \ : \ P_2(A)\}.$$

Since, $\mathscr{E}$ is an algebra, $\mathscr{E} \subset \mathscr{F}_2 \subset \mathscr{F}_1$. Moreover, by similar reasoning $\mathscr{F}_2$ is a monotone class and so we get the result. □

Definition 3.1.11 (π class)**.** *Let X be a non-empty set and let $\mathscr{E}$ be a collection of subsets of X such that:*

(1) For any $A, B \in \mathscr{E}$, $A \cap B \in \mathscr{E}$,

then $\mathscr{E}$ is called a $\pi-$ class.

Definition 3.1.12 (λ class)**.** *Let X be a non-empty set. The collection of subsets of X, $\mathscr{E}$ is called a $\lambda-$ class (system) if,*

(1) X is in $\mathscr{E}$.
(2) $A \in \mathscr{E}$ implies $A^{\mathsf{c}} \in \mathscr{E}$.
(3) If (A_n) is a sequence of pairwise disjoint sets in $\mathscr{E}$, then $\bigcup_{n=1}^{\infty} A_n$ is in $\mathscr{E}$.

It is easy to see that a $\lambda-$ class that is also a $\pi-$ class is a sigma algebra. See 3.1.2. An equivalent definition of the $\lambda-$ class due to Dynkin is given as:

Definition 3.1.13 (D class)**.** *Let X be a non-empty set. The collection of subsets of X, $\mathscr{D}$ is called a $D-$ class if,*

(1) X is in $\mathscr{E}$.
(2) $A, B \in \mathscr{E}$ and $A \supset B$ implies $A - B \in \mathscr{E}$.
(3) If (A_n) is a sequence of pairwise disjoint sets in $\mathscr{E}$, then $\bigcup_{n=1}^{\infty} A_n$ is in $\mathscr{E}$.

It is easy to see that a $D-$ class is equivalent to a $\lambda-$ class. Suppose $\mathscr{A}$ is a D class and consider $A \in \mathscr{A}$. We know that X is in $\mathscr{A}$ and since $X \supset A$, $A^{\mathsf{c}} = X - A$ is in $\mathscr{A}$. Thus a $D-$ class is a $\lambda-$ class. Suppose $\mathscr{A}$ is a λ class, and consider $A, B \in \mathscr{A}$ such that $A \supset B$. Then, $A - B = A \cap B^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$. Since $A \supset B$, A^{c} and B are disjoint and so $A^{\mathsf{c}} \cup B$ is in $\mathscr{A}$ and hence its complement is in $\mathscr{A}$.

If $\mathscr{E} \subset \mathscr{P}(X)$ is non-empty, then the smallest $\lambda-$ class generated by $\mathscr{E}$ is the intersection of all the $\lambda-$ classes containing $\mathscr{E}$ and is denoted by $\mathcal{D}(\mathscr{E})$, where $\mathcal{D}$ stands for Dynkin.

Lemma 3.1.1. *Let X be a non-empty set and let $\mathscr{E} \subset \mathscr{P}(X)$ be a non-empty collection of subsets of X. Then $\mathcal{D}(\mathscr{E}) \subset \sigma(\mathscr{E})$.*

Proof. $\sigma(\mathscr{E})$ is a $\lambda-$ class containing $\mathscr{E}$ and so by minimality contains $\mathcal{D}(\mathscr{E})$. □

Theorem 3.1.8. Let X be a non-empty set and let $\mathscr{E} \subset \mathscr{P}(X)$ be a non-empty collection of subsets of X. If $\mathscr{E}$ is a $\pi-$ class, then $\mathcal{D}(\mathscr{E}) = \sigma(\mathscr{E})$.

Proof. We need to show $\mathcal{D}(\mathscr{E}) \subset \sigma(\mathscr{E})$ and $\sigma(\mathscr{E}) \subset \mathcal{D}(\mathscr{E})$. Since, a sigma-algebra is a $\lambda-$ class we immediately get $\sigma(\mathscr{E}) \supset \mathcal{D}(\mathscr{E})$. For the other direction we will use 3.1.3 along with the ideas in the proof used for the monotone class.

We need to show $\mathcal{D}(\mathscr{E})$ is a sigma algebra containing $\mathscr{E}$ and this amounts to showing that $\mathcal{D}(\mathscr{E})$ is a $\pi-$ class. Thus, consider the sub-collection:

$$\mathscr{F} = \{A \in \mathcal{D}(\mathscr{E}) \ : \ \forall (B \in \mathcal{D})\, A \cap B \in \mathcal{D}(\mathscr{E})\}.$$

By construction $\mathscr{F} \subset \mathcal{D}(\mathscr{E})$ and is a $\pi-$ class. If we show that $\mathscr{F} \supset \mathscr{E}$ and $\mathscr{E}$ is a $\lambda-$ class we are done.

It is clear that X is in $\mathscr{F}$. Let A be in $\mathscr{F}$, and let B be any set in $\mathcal{D}(\mathscr{E})$. Then,

$$A^{\mathsf{c}} \cap B = (A \cup B^{\mathsf{c}})^{\mathsf{c}}.$$

Since, $A \in \mathscr{F}$, $A \in \mathcal{D}(\mathscr{E})$ and since $B \in \mathcal{D}(\mathscr{E})$, $B^{\mathsf{c}} \in \mathcal{D}(\mathscr{E})$. Thus, $(A \cup B^{\mathsf{c}})^{\mathsf{c}} \in \mathcal{D}(\mathscr{E})$. Hence for any $B \in \mathcal{D}(\mathscr{E})$, $A^{\mathsf{c}} \cap B \in \mathscr{F}$ whenever $A \in \mathscr{F}$. Let (A_n) be a sequence of sets in $\mathscr{E}$ that are pairwise disjoint and let B be any set in $\mathcal{D}(\mathscr{E})$. Then,

$$(\bigcup_{n=1}^{\infty} A_n) \cap B = \bigcup_{n=1}^{\infty} (A_n \cap B).$$

Since $\mathscr{F} \subset \mathcal{D}(\mathscr{E})$, $A_n \cap B \in \mathcal{D}\mathscr{E}$ for each n and so the countable union is in $\mathcal{D}(\mathscr{E})$.

To show that $\mathscr{F} \supset \mathscr{E}$, we will construct a family $\mathscr{F}_2$ such that $\mathscr{F} \supset \mathscr{F}_2 \supset \mathscr{E}$ and $\mathscr{F}_2$ is a $\lambda-$ class. Let

$$\mathscr{F}_2 = \{A \in \mathcal{D}(\mathscr{E}) \ : \ \forall\,(B \in \mathscr{E})\, A \cap B \in \mathcal{D}(\mathscr{E})\}\,.$$

Since $\mathscr{E}$ is a $\pi-$ class $\mathscr{E} \subset \mathscr{F}_2$. If we show $\mathscr{F}_2$ is a $\lambda-$ class, it would mean that $\mathscr{F}_2 \supset \mathcal{D}(\mathscr{E})$. But since by construction, $\mathscr{F}_2 \subset \mathscr{F}$ we get $\mathscr{F} \supset \mathcal{D}(\mathscr{E}) \supset \mathscr{E}$. Proving that $\mathscr{F}_2$ is a $\lambda-$ class follows the same reasonings as we showed for $\mathscr{F}$.

□

Theorem 3.1.9 ($\pi - \lambda$ Theorem)**.** If $\mathscr{E}$ is a $\pi-$ class and $\mathscr{L}$ is a $\lambda-$ class, then $\mathscr{E} \subset \mathscr{L}$ implies, $\sigma\,(\mathscr{E}) \subset \mathscr{L}$.

Proof. Consider the minimal $\lambda-$ class containing $\mathscr{E}$, $\mathcal{D}(\mathscr{E})$. By definition, $\mathcal{D}(\mathscr{E}) \subset \mathscr{L}$. We showed that $\sigma\,(\mathscr{E}) = \mathcal{D}(\mathscr{E})$ and thus, $\sigma\,(\mathscr{E}) \subset \mathscr{L}$. □

3.2. Measures

We now consider how to measure the size of the sets in a given σ- algebra on a set X.

Definition 3.2.1 (Measure)**.** *Let X be a set on which there is a σ- algebra $\mathscr{M}$. A measure on the measurable space $(X, \mathscr{M})$ is a function*

$$\mu \ : \mathscr{M} \to [0, \infty]$$

satisfying,

(1) $\mu\,(\emptyset) = 0$,

(2) (σ-additivity) If (E_n) be a sequence of pairwise disjoint sets in $\mathscr{M}$, then

$$\mu\left(\dot{\bigcup}_{i \in \mathbb{Z}^+} E_i\right) = \sum_{i=1}^{\infty} \mu\,(E_i).$$

A *pre-measure* is function that satisfies the above on an *algebra* and not necessarily on a σ- algebra. Note that for a *pre-measure*, additivity is statisfied if the countable union of a sequence of sets is in the algebra. We will denote a pre-measure by μ_0. The triple $(X, \mathscr{M}, \mu)$ is called a *measure space*. A *finite measure* is a measure with $\mu\,(X) < \infty$, and a *probability measure* is a measure with $\mu\,(X) = 1$. Note that for a finite measure, for any $E \in \mathscr{M}$, $\mu\,(E) < \infty$. A measure is said to be σ- finite, if $\mathscr{M}$ contains a sequence $A_1 \subset A_2 \subset A_3 \ldots$ such that $\bigcup\limits_{i \in \mathbb{Z}^+} A_i = X$ and $\mu\,(A_i) < \infty$ for every $i \in \mathbb{Z}^+$. Most measures encountered in practice are at least σ- finite, and non-σ- finite measures have some strange behaviour. A measure μ is σ semi-finite if any set $F \in \mathscr{M}$ such that $\mu\,(F) = \infty$ there is a set $E \in \mathscr{M}$ and $E \subset F$ such that $\mu\,(E) < \infty$.

Definition 3.2.2 (Finitely additive measure)**.** *If* (E_i) *are disjoint sets in* $\mathscr{M}$*, and* $\mu\left(\dot{\bigcup}_{i=1}^{n} E\right) = \sum_{i=1}^{n} \mu(E_i)$*, then* μ *is finitely additive measure.*

Note that a pre-measure will always satisfy finite additivity.

Example 3.2.1. *Let* X *be an uncountable set and let* $\mathscr{M}$ *be the* σ*- algebra,*

$$\{A \subset X \ : \ A \, \textit{is countable or} \, A^{\mathsf{c}} \textit{ is countable}\}$$

Then the following set function defined on $\mathscr{M}$*,*

$$\mu(E) = \begin{cases} 1 & E \textit{ is countable} \\ 0 & E^{\mathsf{c}} \textit{ is countable} \end{cases}$$

is a finitely additive measure but is not countably additive.

We collect all the important properties of a measure in the following theorem. The following properties will also be satisfied by a pre-measure provided we check that the countable union and intersection of a sequence of sets belong to the algebra.

Theorem 3.2.1 (Properties of measure)**.** Let $(X, \mathscr{M}, \mu)$ be a measure space. Then,

Property 1 (finite-additivity) $E, F \in \mathscr{M}$ and $E \cap F = \emptyset$, then

$$\mu\left(E \dot{\bigcup} F\right) = \mu(E) + \mu(F).$$

Property 2 (monotonicity) If $E, F \in \mathscr{M}$ and $E \subset F$ then

$$\mu(E) \leq \mu(F).$$

Moreover if $\mu(F) < \infty$, then

$$\mu(F - E) = \mu(F) - \mu(E).$$

Property 3 (sub-additivity) If (E_i) be a sequence of sets in $\mathscr{M}$, then

$$\mu\left(\bigcup_{i \in \mathbb{Z}^+} E_i\right) \leq \sum_{i=1}^{\infty} \mu(E_i).$$

Property 4 (continuity from below) If $(E_i) \in \mathscr{M}$, and

$E_1 \subset E_2 \subset E_3 \ldots$, then

$$\mu\left(\bigcup_{i \in \mathbb{Z}^+} E_i\right) = \lim_{i \to \infty} \mu(E_i).$$

Property 5 (continuity from above) If $(E_i) \in \mathscr{M}$, and

$$E_1 \supset E_2 \supset E_3 \ldots$$

and $\mu(E_n) < \infty$ for some $n \in \mathbb{Z}^+$, then

$$\mu\left(\bigcap_{i \in \mathbb{Z}^+} E_i\right) = \lim_{i \to \infty} \mu(E_i).$$

Proof. We prove in order.

(1) Let $E_1 = E$ and $E_2 = F$ and $E_i = \emptyset$ for all $i \geq 3, i \in \mathbb{Z}^+$. Then, $\mu\left(E \dot{\bigcup} F\right) = \mu\left(\dot{\bigcup}_{i \in \mathbb{Z}^+} E_i\right) = \mu(E) + \mu(F)$.

(2) $F = E \dot{\bigcup} F \cap E^{\mathsf{c}}$, and so $\mu(F) = \mu(E) + \mu(F \cap E^{\mathsf{c}})$, and hence $\mu(E) \leq \mu(F)$. (Since $E, F \in \mathscr{M}$, $F \cap E^{\mathsf{c}}$ is also in $\mathscr{M}$, the result follows from (1) because μ is a non-negative set function). Now if $\mu(F) < \infty$, then $\mu(E) < \infty$ and so we can subtract the $\mu(E)$ to obtain $\mu(F) - \mu(E) = \mu(F \cap E^{\mathsf{c}}) = \mu(F - E)$. Note that, we actually just need the finiteness of $\mu(E)$.

(3) From 3.1.1, we can construct disjoint sets out of regular sets as follows, let $F_1 = E_1$ and $F_k = E_k - \bigcup_{i=1}^{k-1} E_i$ for $k \geq 2$. Note that each $\mu(F_k) \leq \mu(E_k)$ and $\bigcup_{i\in\mathbb{Z}^+} E_i = \dot{\bigcup}_{i\in\mathbb{Z}^+} F_i$. Hence,

$$\mu\left(\bigcup_{i\in\mathbb{Z}^+} E_i\right) = \mu\left(\dot{\bigcup}_{i\in\mathbb{Z}^+} F_i\right) = \sum_{i=1}^{\infty} \mu(F_i) \leq \sum_{i=1}^{\infty} \mu(E_i).$$

(4) From 3.1.1, we can construct disjoint sets as above, let $F_1 = E_1$ and $F_k = E_k - \bigcup_{i=1}^{k-1} E_i = E_k - E_{k-1}$ for $k \geq 2$. See 3.2. Also note that $E_n = \dot{\bigcup}_{i=1}^{n} F_i$ and so $\mu(E_n) = \mu\left(\dot{\bigcup}_{i=1}^{n} F\right) = \sum_{i=1}^{n} \mu(F_i)$. Thus,

$$\begin{aligned}
\mu\left(\bigcup_{i\in\mathbb{Z}^+} E_i\right) &= \mu\left(\dot{\bigcup}_{i\in\mathbb{Z}^+} F_i\right) \\
&= \sum_{i=1}^{\infty} \mu(F_i) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(F_i) \\
&= \lim_{n\to\infty} \mu(E_n).
\end{aligned}$$

(5) Let $F_k = E_n - E_k$ for $k > n$, and $F_k = \emptyset$ for $k \leq n$. Note that $E_n \supset E_{n+1} \supset E_{n+2} \ldots$, and so $F_{n+1} \subset F_{n+2} \ldots$ i.e. $F_{n+1} = E_n - E_{n+1} \subset E_n - E_{n+2} = F_{n+2}$ and so on. See 3.3. Hence,

$$\bigcup_{i\geq n+1} F_i = \bigcup_{i\geq n+1} (E_n - E_i) = E_n - \bigcap_{i\geq n+1} E_i.$$

Thus we get,

$$\mu\left(\bigcup_{i\in\mathbb{Z}^+} F_i\right) = \mu(E_n) - \mu\left(\bigcap_{i\geq n+1} E_i\right).$$

because, $E_{n+k} \subset E_n$ and $\mu(E_n) < \infty$. Also we get a increasing sequence of sets $F_i \subset F_{i+1} \subset F_{i+2} \ldots$, for $i > n$. Thus we know that,

$$\mu\left(\bigcup_{i\in\mathbb{Z}^+} F_i\right) = \lim_{i\to\infty} \mu(F_i) = \lim_{k\to\infty} \mu(E_n - E_{n+k}).$$

Since $E_{n+k} \subset E_n$ and $\mu(E_n) < \infty$ we get,

$$\mu\left(\bigcup_{i\in\mathbb{Z}^+} F_i\right) = \lim_{i\to\infty} \mu(F_i) = \mu(E_n) - \lim_{k\to\infty} \mu(E_{n+k}).$$

Thus equating the two we get our result.

□

The next theorem says that in a sigma-finite measurable space, there cannot be a countable collection of disjoint sets with positive measure. The proof of the theorem is highly illustrative of a common technique that pops up in measure theory.

Theorem 3.2.2. Let $(X, \mathscr{M}, \mu)$ be a sigma-finite measure space. Then, $\mathscr{M}$ cannot contain an uncountable, disjoint collection of sets of positive measure.

Proof. Let $\mathscr{E} = \{E_\alpha \,:\, \alpha \in A\}$ be the collection of subsets of X such that for each $\alpha \in A$, $\mu(E_\alpha) > 0$. We need to show that $\mathscr{E}$ is a countable. Since, we have a sigma finite space, there is a sequence of sets (X_n) such that $X_n \nearrow X$ and $\mu(X_n) < \infty$ for each n. For any $E \in \mathscr{E}$,

$$E \subset X = \bigcup_{n=1}^{\infty} X_n,$$

thus, $E = \bigcup\limits_{n=1}^{\infty} (E \cap X_n)$. For any $\epsilon > 0$, consider the collection,

$$\mathscr{E}_{n,\epsilon} = \{E \in \mathscr{E} \,:\, \mu((E \cap X_n)) > \epsilon\}.$$

If E is any set in $\mathscr{E}$, then there must be an n, ϵ such that $E \in \mathscr{E}_{n,\epsilon}$. Using the archimedes principle then there must be a $k \in \mathbb{Z}^+$ such that $\frac{1}{k} < \epsilon$. Hence, let us consider the collection,

$$\mathscr{E}_{n,k} = \left\{E \in \mathscr{E} \,:\, \mu((E \cap X_n)) > \frac{1}{k}\right\}.$$

Clearly for each n, k, $\mathscr{E}_{n,k} \subset \mathscr{E}$. Also, by reasoning above, for any $E \in \mathscr{E}$ there must be a $k, n \in \mathbb{Z}^+$, such that $E \in \mathscr{E}_{n,k}$. Hence,

$$\mathscr{E} = \bigcup_{n,k=1}^{\infty} \mathscr{E}_{n,k}.$$

Hence, if we show that each $\mathscr{E}_{n,k}$ is countable, then we are done. Consider a sequence of finite sets $E_1, E_2, \ldots, E_m$ in one such $\mathscr{E}_{n,k}$. Since, these sets are

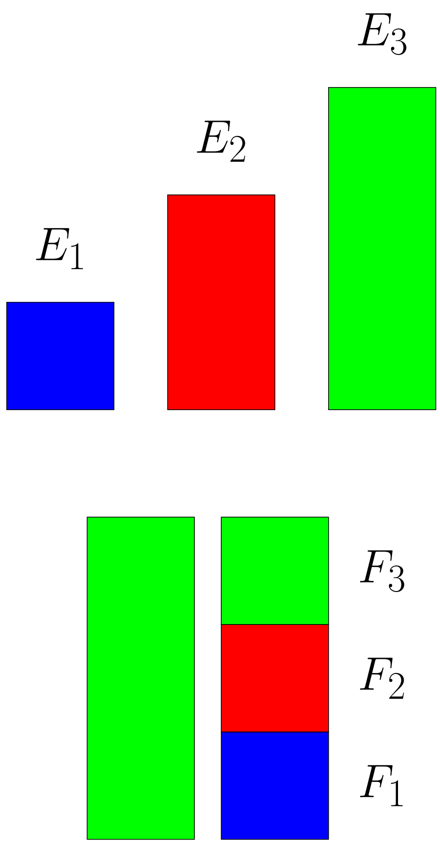

Figure 3.2. Illustration of proof 3.2.1 (4)

disjoint,

$$\frac{m}{k} \leq \sum_{i=1}^{m} \mu\left(E_i \cap X_n\right) = \mu\left(\bigcup_{i=1}^{n} \left(E_i \cap X_n\right)\right) \leq \mu\left(X_n\right).$$

Thus, $m \leq k\mu\left(X_n\right)$. Hence we only have a finite choice for the index m. Thus, $\mathscr{E}_{n,k}$ is finite. □

3.3. Sets of measure zero and completion of measure

We know that we want to deal with sets of measure zero. There is a technical issue about such sets that we settle now.

Definition 3.3.1 (Sets of measure zero)**.** *If* $(X, \mathscr{M}, \mu)$ *is a measure space, a set* $E \in \mathscr{M}$ *with* $\mu(E) = 0$ *is called a set of* measure zero.

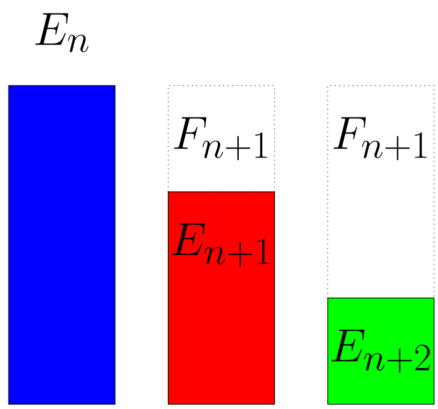

Figure 3.3. Illustration of proof 3.2.1 (5)

Definition 3.3.2 (Almost everywhere)**.** *If a statement about points $x \in X$ is true except for x in a set of* measure zero*, we say that the statement is true* almost everywhere *and denote it by a.e.*

Proposition 3.3.1 (Countable union of sets of measure 0)**.** *A countable union of sets of measure zero has measure zero.*

Proof. From the sub-additivity property of measure, if (E_i) is a sequence of measure zero then $\mu\left(\bigcup_{i\in\mathbb{Z}^+} E_i\right) \leq \sum_{i=1}^{\infty} \mu(E_i) = 0$. Since measure is a non-negative set function we get $\mu\left(\bigcup_{i\in\mathbb{Z}^+} E_i\right) = 0$. □

Remark 3.3.1. *If $\mu(F) = 0$ for some $F \in \mathscr{M}$, then $\mu(E) = 0$ for any $E \subset F$ whenever $E \in \mathscr{M}$. However, E may not be in $\mathscr{M}$. This is a technical issue that we need to resolve. We will do so my adding to the $\mathscr{M}$ all those subsets of sets in $\mathscr{M}$ that have measure* 0*. This is called the completion of the measure μ.*

Definition 3.3.3 (Complete measure space)**.** *If $(X, \mathscr{M}, \mu)$ is a measure space such that $\mathscr{M}$ contains all subsets of sets in $\mathscr{M}$ with measure* 0*, then $(X, \mathscr{M}, \mu)$ is* complete.

Completeness eliminates some annoying issues and it can always be obtained by enlarging the domain of a given measure to obtain an equivalent measure in the following sense:

Theorem 3.3.1 (Completion of a measure)**.** Let $(X, \mathscr{M}, \mu)$ be a measure space. Let $\mathscr{N} = \{N \in \mathscr{M} \ : \ \mu(N) = 0\}$. Define,

$$\overline{\mathscr{M}} = \{E \cup F \ : \ E \in \mathscr{M}, \exists (N \in \mathscr{N}) \text{ such that } F \subset N\},$$

then $\overline{\mathscr{M}}$ is a σ- algebra on X that contains $\mathscr{M}$. Moreover, the unique measure $\bar{\mu}$ on $\overline{\mathscr{M}}$ defined by $\bar{\mu}(E \cup F) = \mu(E)$ for all $E \in \mathscr{M}$ makes $\left(X, \overline{\mathscr{M}}, \bar{\mu}\right)$ complete.

Proof. Clearly $\mathscr{M} \subset \overline{\mathscr{M}}$. We need to show that $\overline{\mathscr{M}}$ is a σ- algebra. Let $A \in \overline{\mathscr{M}}$. Then $A = E \cup F$ such that there is a $N \in \mathscr{N}$ with $F \subset N$. Note that $F \subset N \implies N^{\mathsf{c}} \subset F^{\mathsf{c}}$ and so $F^{\mathsf{c}} = N^{\mathsf{c}} \dot{\bigcup} N - F$. Thus $A^{\mathsf{c}} = E^{\mathsf{c}} \cap F^{\mathsf{c}} = E^{\mathsf{c}} \cap (N^{\mathsf{c}} \cup (N - F))$. Thus the distributive law yields,

$$A^{\mathsf{c}} = (E^{\mathsf{c}} \cap N^{\mathsf{c}}) \cup (E^{\mathsf{c}} \cap N \cap F^{\mathsf{c}}).$$

Now $(E^{\mathsf{c}} \cap N^{\mathsf{c}}) \in \mathscr{M}$ since both $E, N \in \mathscr{M}$. Also $(E^{\mathsf{c}} \cap N \cap F^{\mathsf{c}}) \subset N$ and hence $A^{\mathsf{c}} \in \overline{\mathscr{M}}$. Let (A_i) be a sequence of sets in $\mathscr{M}$. Then each $A_i = E_i \cup F_i$ such that there is a $N_i \in \mathscr{N}$ such that $F_i \subset N_i$. Thus,

$$\bigcup_{i \in \mathbb{Z}^+} A_i = (\bigcup_{i \in \mathbb{Z}^+} E_i) \bigcup (\bigcup_{i \in \mathbb{Z}^+} F_i).$$

$\bigcup_{i \in \mathbb{Z}^+} E_i \in \mathscr{M}$ and $\bigcup_{i \in \mathbb{Z}^+} F_i \subset \bigcup_{i \in \mathbb{Z}^+} N_i$. From 3.3.1, $\bigcup_{i \in \mathbb{Z}^+} N_i \in \mathscr{N}$ and hence $\bigcup_{i \in \mathbb{Z}^+} A_i \in \overline{\mathscr{M}}$. Thus $\overline{\mathscr{M}}$ is a σ- algebra.

We have to check if $\bar{\mu}(A)$ is well defined i.e if $A = E_1 \cup F_1 = E_2 \cup F_2$ then $\mu(E_1) = \mu(E_2)$. To see this, $E_1 \subset E_1 \cup F_1 = E_2 \cup F_2 \subset E_2 \cup N_2 \in \mathscr{M}$. Thus $\mu(E_1) \leq \mu(E_2) + \mu(N_2) = \mu(E_2)$. Similarly, $\mu(E_2) \leq \mu(E_1)$.

Thus to show that $\left(X, \overline{\mathscr{M}}, \bar{\mu}\right)$ is complete we need to show that for any $A \in \overline{\mathscr{M}}$, if $\bar{\mu}(A) = 0$, then for any $B \subset A$ we must have $B \in \overline{\mathscr{A}}$. Since $A = E \cup F$ where $F \subset N$ for some $N \in \mathscr{N}$, we have $\bar{\mu}(A) = \mu(E)$. But this implies that $\mu(E) = 0$ and so $E \in \mathscr{N}$. Thus $A \in \mathscr{N}$. But $B \subset A$ and so $B = \emptyset \cup B$ where $\emptyset \in \mathscr{M}$ and $B \subset A \in \mathscr{N}$. Thus $B \in \overline{\mathscr{M}}$. $\square$

Example 3.3.1. *Consider a set $X \neq \emptyset$ and the minimal sigma algebra $\mathscr{M} = \{X, \emptyset\}$ with a measure μ defined to be 0 for any set in the $\mathscr{M}$. Clearly $(X, \mathscr{M}, \mu)$ is not complete. To get a completion of X we must add all subsets of X to the σ- algebra. Thus $\overline{\mathscr{M}} = \mathscr{P}(X)$.*

3.4. Construction of measures

A σ- algebra is a huge set and specifying a set function that satisfies the requirements of a measure is not a trivial pursuit. Instead we work with set functions that can be easily constructed and then derive a measure from it by restricting or extending its domain. One such way is the *outer measure*.

As we will see an outer measure can be derived easily from certain set functions. Once an outer measure is constructed it can be restricted to generate a measure.

Definition 3.4.1 (Outer measure)**.** *If X is a non-empty set, an outer measure on X is a function $\mu^* : \mathscr{P}(X) \to [0, \infty]$ such that,*

(1)

$$\mu^*(\emptyset) = 0.$$

(2) (Monotonicity) For any $A \subset B$,

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

(3) (sub-additivity) For any (A_i) in $\mathscr{P}(X)$,

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

It is easy to see that every measure whose domain is the $\mathscr{P}(X)$ is an outer measure. Now we will show that an outer measure can be restricted to yield a measure. We will use Carathédory's theorem to establish this result.

Definition 3.4.2 (μ^*- measurable sets: Carathédory's Criterion)**.** *Given any outer measure μ^*, we say that a set E is μ^*- measurable if,*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}}),$$

for any subset $A \subset X$.

We want those sets E such that no matter what $A \subset X$ we take, outer measure of A is additive on A (w.r.t E). This is shown in 3.4. Note that A itself is not required to be μ^*- measurable. A vague motivation for this is, if E is a *good* set and $A \supset E$, Carathédory's criteria states that its outer measure $\mu^*(E) = \mu^*(E \cap A)$ is equal to $\mu^*(A) - \mu^*(A \cap E^{\mathsf{c}})$. While the later concerns measuring E from *inside* A, the term $\mu^*(E) = \mu^*(E \cap A)$ concerns measuring E from *outside* E. This is seen from the right side of 3.4 and thus *good* sets have the same *measure*—inside or outside.

We can just say measurable instead of μ^*- measurable if μ^* is clear from the context.

Proposition 3.4.1. *Given any outer measure μ^*, we say that a set E is μ^*- measurable if,*

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}}),$$

for any subset $A \subset X$.

Proof. From the 3.4 it is easy to see that $A = A \cap E \dot{\bigcup} A \cap E^{\mathsf{c}}$. Since μ^* is sub-additive we have $\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}})$. Thus, for E to be

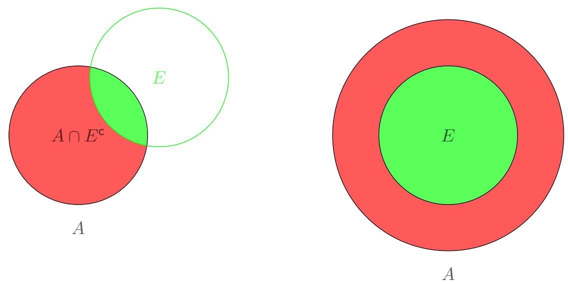

Figure 3.4. Carathéodory criteria: $\mu^*(A)$ is the sum of the two colors.

μ^* - measurable, we must have,

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}}),$$

for any subset $A \subset X$. $\square$

Theorem 3.4.1 (Carathéodory Theorem)**.** Let X be a non-empty set with an outer measure μ^* defined on $\mathscr{P}(X)$. Define,

$$\mathscr{M} = \{E \subset X \ : \ E \text{ is } \mu^* \text{- measurable}\}.$$

Then, $\mathscr{M}$ is a σ- algebra and the restriction of μ^* on $\mathscr{M}$ is a measure. Thus, $\left(X, \mathscr{M}, \mu^*\big|_{\mathscr{M}}\right)$ is a measurable space.

Proof. We first need to show that $\mathscr{M}$ is a σ- algebra. A priori we don't know if $\mathscr{M}$ is empty or not. To this end we will show that $\emptyset \in \mathscr{M}$. If $\mu^*(E) = 0$ then, we need to show that for any $A \subset X$, $\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}})$. Since, $A \supset A \cap E^{\mathsf{c}}$ and hence $\mu^*(A) \geq A \cap E^{\mathsf{c}}$ from monotonicity of μ^*. Similarly, $E \supset A \cap E$ and $\mu^*(E) \geq A \cap E$. Thus, adding, $\mu^*(A) + \mu^*(E) \geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}})$. Since $\mu^*(E) = 0$, the result follows. Thus E is μ^* - measurable. This means that $\emptyset \in \mathscr{M}$ since $\mu^*(\emptyset) = 0$.

If $E \in \mathscr{M}$ then for any $A \subset X$, $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}})$. Replacing E by E^{c} we get the same expression and thus $E^{\mathsf{c}} \in \mathscr{M}$.

Let $E_1, E_2 \in \mathscr{M}$. Then they satisfy Carathéodory criteria 3.4.2. Thus,

$$\mu^*(A_1) = \mu^*(A_1 \cap E_1) + \mu^*(A_1 \cap E_1{}^{\mathsf{c}})$$
$$\mu^*(A_2) = \mu^*(A_2 \cap E_2) + \mu^*(A_2 \cap E_2{}^{\mathsf{c}})$$

for any $A_1, A_2 \subset X$. Pick $A_1 = A \subset X$ and $A_2 = A \cap E_1$ and so we get,

$$\mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A \cap E_1{}^{\mathsf{c}})$$
$$\mu^*(A \cap E_1{}^{\mathsf{c}}) = \mu^*(A \cap E_1{}^{\mathsf{c}} \cap E_2) + \mu^*(A \cap E_1{}^{\mathsf{c}} \cap E_2{}^{\mathsf{c}})$$

Substituting for $\mu^*(A \cap E_1^{\mathsf{c}})$ we get,

$$\mu^*(A) = \mu^*(A \cap E_1) + \mu^*(A \cap E_1{}^{\mathsf{c}} \cap E_2) + \mu^*(A \cap (E_1 \cup E_2)^{\mathsf{c}}).$$

Now, $A \cap (E_1 \cup E_2) = (A \cap E_1) \dot{\bigcup} (A \cap E_1{}^{\mathsf{c}} \cap E_2)$ is easily seen as the disjoint union of two sets X, Y where $X = A \cap E_1$, $Y = A \cap E_2$, and hence,

$$\mu^*(A \cap E_1) + \mu^*(A \cap E_1{}^{\mathsf{c}} \cap E_2) \geq \mu^*(A \cap (E_1 \cup E_2)).$$

Thus,

$$\mu^*(A) \geq \mu^*(A \cap (E_1 \cup E_2)) + \mu^*(A \cap (E_1 \cup E_2)^{\mathsf{c}})$$

which from 3.4.1 means that $E_1 \cup E_2 \in \mathscr{M}$. Thus we get finite additivity. Hence we have shown that $\mathscr{M}$ is an algebra.

Showing that $\mathscr{M}$ is closed under a countable union of an arbitrary sequence of sets is a little tricky. Instead we will show that $\mathscr{M}$ is closed under countable union of a sequence of disjoint sets. Then $\mathscr{M}$ will be a σ- algebra by 3.1.2. We will do this in 3 steps: Let (E_i) be a sequence of pairwise *disjoint* sets of $\mathscr{M}$ and let $S_n = \bigcup_{i=1}^{n} E_i$ and $S = \bigcup_{i \in \mathbb{Z}^+} E_i$.

Step 1: We will show that for any $n \geq 1$,

$$\mu^*(A \cap S_n) = \sum_{i=1}^{n} \mu^*(A \cap E_i).$$

Note that since we proved $\mathscr{M}$ is an algebra, $\mathscr{M}$ is closed under finite unions and hence $S_n \in \mathscr{M}$. We will use induction on n. The case $n = 1$ is trivial. Suppose it is true for some $k > 1 \in \mathbb{Z}^+$, then since $S_k \in \mathscr{M}$ we get,

$$\begin{aligned}
\mu^*(A \cap S_{k+1}) &= \mu^*(A \cap S_{k+1} \cap S_k) + \mu^*(A \cap S_{k+1} \cap S_k^{\mathsf{c}}) \\
&= \mu^*(A \cap S_k) + \mu^*(A \cap E_{k+1}) \\
&= \sum_{i=1}^{k} \mu^*(A \cap E_i) + \mu^*(A \cap E_{k+1}) \\
&= \sum_{i=1}^{k+1} \mu^*(A \cap E_i)
\end{aligned}$$

Step 2: We will show,

$$\mu^*(A \cap S) = \sum_{i=1}^{\infty} \mu^*(A \cap E_i).$$

From sub-additivity of μ^*,

$$\mu^*(A \cap S) = \mu^*\left(\bigcup_{i \in \mathbb{Z}^+} E_i\right) \leq \sum_{i=1}^{\infty} \mu^*(E_i)$$

To show the other direction, $S \supset S_n$ for any $n \in \mathbb{Z}^+$ and hence from monotonicity and **Step 1**,

$$\mu^*(A \cap S) \geq \mu^*(A \cap S_n) = \sum_{i=1}^{n} \mu(E_i).$$

Taking the limit as $n \to \infty$ we get the desired result.

Step 3: We will show,

$$\mu^*(A) \geq \mu^*(A \cap S) + \mu^*(A \cap S^{\mathsf{c}})$$

Consider an $n \in \mathbb{Z}^+$ and since $S_n \in \mathscr{M}$ we have,

$$\mu^*(A) \geq \mu^*(A \cap S_n) + \mu^*(A \cap S_n^{\mathsf{c}})$$

But for any $n \in \mathbb{Z}^+$, $S_n^{\mathsf{c}} \supset S^{\mathsf{c}}$ and thus using **Step 1**

$$\begin{aligned} \mu^*(A) \geq & \mu^*(A \cap S_n) + \mu^*(A \cap S^{\mathsf{c}}) \\ = & \sum_{i=1}^{n} \mu^*(E_i) + \mu^*(A \cap S^{\mathsf{c}}) \end{aligned}$$

taking the limit as $n \to \infty$ and using **Step 2** we get the result.

Thus $\mathscr{M}$ is a σ- algebra. It is easy to show that $\mu = \mu^*\big|_{\mathscr{M}}$ is a measure. Indeed all we need to show is that μ satisfies σ- additivity. Let (E_i) be a sequence of pairwise *disjoint* sets of $\mathscr{M}$ and let $S = \bigcup_{i \in \mathbb{Z}^+} E_i$. From **Step 2**,

$$\mu^*(A \cap S) = \sum_{i=1}^{\infty} \mu^*(A \cap E_i)$$

for any $A \subset X$. If we restrict μ^* to $\mathscr{M}$, this means,

$$\mu(A \cap S) = \sum_{i=1}^{\infty} \mu(A \cap E_i)$$

for any $A \in \mathscr{M}$. Since $S \in \mathscr{M}$ take $A = S$ and thus

$$\mu\left(\dot{\bigcup}_{i \in \mathbb{Z}^+} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

Table 3.1. Properties of measure, pre-measure and outer-measure

Properties	Measure μ	Pre-Measure μ_0	Outer-Measure μ^*
domain	σ- algebra	algebra	$\mathscr{P}(X)$
monotonicity	✓	✓	✓
finite-additvity	✓	✓	-
finite-sub-additvity	✓	✓	✓
countable-additvity	✓	-	-
countable-sub-additvity	✓	-	✓

Also interersting to note is that $(X, \mathscr{M}, \mu)$ is complete. To see this take any $E \in \mathscr{M}$ such that $\mu(E) = 0$ and let $B \subset E$. Clearly $B \subset X$ and thus,

$$\mu^*(B) \leq \mu^*(E) = \mu(E) = 0.$$

Hence, $\mu^*(B) = 0$ which means $B \in \mathscr{M}$ (as proved earlier). □

Thus we have seen that any outer measure can be restricted to a measure. Hence for this Theorem to be useful we must have an outer measure. We noted that describing a measure for all members of a σ- algebra is no easy matter but now we have this outer measure on the maximal σ- algebra! This is not a concern since the next theorem shows that we can easily construct an outer measure from simple set functions defined on a simpler class. For example, lebesgue measure is defined on intervals. At this point it will do us well to tabulate the properties of all the measures we have seen. This is done in 3.1.

Example 3.4.1. *Let X be an infinite set and define,*

$$\mu^*(E) = \begin{cases} |E| & \text{if } E \text{ is finite,} \\ \infty & \text{if } E \text{ is infinite.} \end{cases}$$

Then μ^* is an outer measure.

Proof. Clearly $\mu^* : \mathscr{P}(X) \to [0, \infty]$ and $\mu^*(\emptyset) = 0$. Let us check monotonicity. If $A \subset B \subset X$, then if both are finite we have $|A| \leq |B|$ and so $\mu^*(A) \leq \mu^*(B)$. If B is infinite then its obvious. Consider a sequence (A_i) of sets in X. If any of the A_i is infinite then $\mu^*\left(\bigcup\limits_{i \in \mathbb{Z}^+} A_i\right) = \infty = \sum\limits_{i=1}^{\infty} \mu^*(A_i)$. Here we treat terms like $\infty + \infty$ to be equal to ∞. If all A_i are finite then there are two cases. Either $\bigcup\limits_{i \in \mathbb{Z}^+} A_i$ is finite or it is infinite. (For expample, $|A_i| = i$, then $\bigcup\limits_{i \in \mathbb{Z}^+} A_i$ is infinite if A_i are pairwise disjoint). If $\bigcup\limits_{i \in \mathbb{Z}^+} A_i$ is finite then we know that $|\bigcup_{i\in\mathbb{Z}^+} A_i| \leq \sum_{i=1}^{\infty} |A_i|$ and hence the result follows. If $\bigcup_{i\in\mathbb{Z}^+} A_i$ is infinite then we can create a sequence of disjoint finte sets $F_k = A_k - \bigcup_{i=1}^{k-1} A_i$. Then, $\bigcup_{i\in\mathbb{Z}^+} A_i = \dot{\bigcup}_{i\in\mathbb{Z}^+} F_i$. Thus $\mu^*\left(\bigcup_{i\in\mathbb{Z}^+} F_i\right) = \sum_{i=1}^{\infty} \mu^*(F_i) = \infty$ since we have finite disjoint sets, but $\sum_{i=1}^{\infty} \mu^*(F_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$ and hence we get, $\mu^*\left(\bigcup_{i\in\mathbb{Z}^+} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$. All finite sets are μ^* measurable. □

In 3.5, we illustrate a way of constructing outer measure. Say we have a set A_k. If we want to define an outer measure we start with a sequence of simple sets E_i^k whose union $\bigcup_i E_i^k \supset A_k$. If we have a simple set function (maybe a pre-measure) ρ for these sets then we get an overestimated outer measure of A given by $\sum_{i=1}^{\infty} \rho(E_i^k)$. The idea is now to take all those collections of sets whose union contain A_k and then take the infimum of these overestimated outer measure. The infimum is then precisely the outer measure of A_k.

Theorem 3.4.2 (Constructing outer measure)**.** Let X be a non-empty set, $\mathscr{E} \subset \mathscr{P}(X)$ be non-empty family of subsets of X such that $\emptyset, X \in \mathscr{E}$. Let there be a function $\rho : \mathscr{E} \to [0,\infty]$ such that $\rho(\emptyset) = 0$. For any $A \in \mathscr{P}(X)$, define

$$\mu^*(A) = \inf\left\{\sum_{i=1}^{\infty} \rho(E_i) \ : \ \{E_i\} \subset \mathscr{E} \text{ and } A \subset \bigcup_{i\in\mathbb{Z}^+} E_i\right\},$$

then, $\mu^*(A)$ is an outer measure (induced by ρ).

Proof. We will first check that the infimum exists. The set is certainly not empty since $X \in \mathscr{E}$ and $A \subset X$; atleast one such $E_i = X$ exists. Morever we are taking infimum over positive reals that are bounded below by 0 and hence the infimum exists.

Lets check if $\mu^*(\emptyset) = 0$. Since $\emptyset \in \mathscr{E}$ it is trivally contained in a collection of empty sets. Since $\rho(\emptyset) = 0$ we get the result.

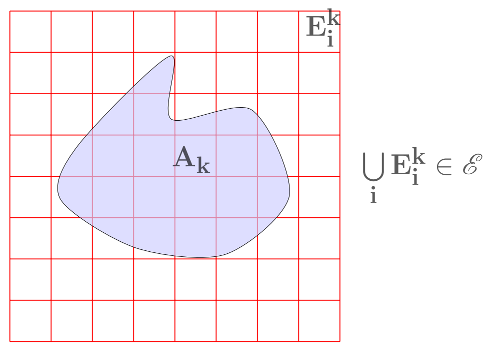

Figure 3.5. Construction of outer measure

Let us see if we get monotonicity. Let $A \subset B$. For any collection $\{E_i\} \subset \mathscr{E}$ if $\bigcup_{i\in\mathbb{Z}^+} E_i \supset B$ then $\bigcup_{i\in\mathbb{Z}^+} E_i \supset A$. Thus the set,

$$\left\{\sum_{i=1}^{\infty} \rho(E_i) \ : \ \{E_i\} \subset \mathscr{E} \text{ and } A \subset \bigcup_{i\in\mathbb{Z}^+} E_i\right\}$$
$$\supset$$
$$\left\{\sum_{i=1}^{\infty} \rho(E_i) \ : \ \{E_i\} \subset \mathscr{E} \text{ and } B \subset \bigcup_{i\in\mathbb{Z}^+} E_i\right\},$$

hence the result follows.

Let (A_k) be a sequence of sets in X such that for each k there is a sequence of sets $\left(E_i^k\right)$ such that,

$$A_k \subset \bigcup_i E_i^k \text{ and,}$$
$$\sum_{i=1}^{\infty} \rho(E_i^k) \leq \mu^*(A_k) + \epsilon,$$

for any $\epsilon > 0$. Pick $\epsilon = \frac{1}{2^k}$. Now,

$$\bigcup_k A_k \subset \bigcup_{i,k} E_i^k.$$

Let $A = \bigcup_k A_k$, thus $A \subset \bigcup_{i,k} E_k^i$, and so by definition

$$\begin{aligned}\mu^*(A) &\le \sum_{i,k=1}^{\infty} \rho(E_k^i)\\ &\le \sum_{k=1}^{\infty}(\mu^*(A_k) + \frac{1}{2^k})\\ &\le \sum_{k=1}^{\infty} \mu^*(A_k).\end{aligned}$$

Thus μ^* is an outer measure. □

Note, at this point there is no relation between ρ and the outer measure induced by ρ. Very little demands are made on ρ and its domain $\mathscr{E}$. An interesting question to ask would be—what if we had some additional structure on $\mathscr{E}$, particularly what if $\mathscr{E}$ were an algebra? In that case, if additionaly $\rho = \mu_0$ is a pre-measure, we have a relation between μ_0 and the outer measure μ^* induced by μ_0. The following Proposition showcases this relation.

Proposition 3.4.2. *If μ_0 is a pre measure on $\mathscr{E}$ and the outer measure induced by μ_0 is given as in 3.4.2, then*

(1) $\mu^*\big|_{\mathscr{E}} = \mu_0$.

(2) Every set in $\mathscr{E}$ is μ^*- measurable.

Proof. We prove in order.

(1) We need to show that $\mu^*\big|_{\mathscr{E}} \le \mu_0$ and $\mu_0 \le \mu^*\big|_{\mathscr{E}}$. Let $A \in \mathscr{E}$. Clearly, $A \subset A$, consider the collection $\{E_i\}$ where $E_1 = A$ and $E_i = \emptyset, i \ge 2$. Hence $\mu^*\big|_{\mathscr{E}}(A) \le \mu_0(A)$. To prove the other inequality, for any $\epsilon > 0$ there is a collection $\{E_i\} \subset \mathscr{E}$ such that $A \subset \bigcup_{i \in \mathbb{Z}^+} E_i$ and,

$$\sum_{i=1}^{\infty} \mu_0(E_i) < \mu^*\big|_{\mathscr{E}}(A) + \epsilon.$$

Since μ_0 is a pre-measure and $A \in \mathscr{E}$, if $\bigcup_{i \in \mathbb{Z}^+} E_i \in \mathscr{E}$ we get from sub-additivity of μ_0,

$$\mu_0(A) \le \sum_{i=1}^{\infty} \mu_0(E_i).$$

However, it may happen that $\bigcup\limits_{i\in\mathbb{Z}^+} E_i \notin \mathscr{E}$. In that case, observe that,

$$A = A \cap \bigcup_{i\in\mathbb{Z}^+} E_i = \bigcup_{i\in\mathbb{Z}^+} A \cap E_i.$$

We will write union of $A \cap E_i$ as a disjoint union using 3.1.1. Let $X_j = A \cap E_j$, $F_j = X_j - \bigcup\limits_{i=1}^{j} X_i, j \geq 2, F_1 = X_1$. Each $F_j \in \mathscr{E}$, moreover $\dot{\bigcup}_j F_j = A \in \mathscr{E}$, thus

$$\mu_0(A) = \sum_{i=1}^{\infty} \mu_0(F_i) \leq \sum_{i=1}^{\infty} \mu_0(E_i).$$

Thus,

$$\mu_0(A) \leq \sum_{i=1}^{\infty} \mu_0(E_i).$$

Hence,

$$\mu_0(A) < \mu^*\big|_{\mathscr{E}}(A) + \epsilon,$$

since ϵ is arbitrary we get the result.

(2) Let $E \in \mathscr{E}$. To show that E is μ^*- measurable we have to show,

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}})$$

for any $A \subset X$. Fix an $\epsilon > 0$, then there is a collection $\{B_i\} \subset \mathscr{E}$ such that $\bigcup\limits_{i=1}^{\infty} B_i \supset A$ and,

$$\mu^*(A) + \epsilon > \sum_{i=1}^{\infty} \mu_0(B_i).$$

Since $E \in \mathscr{E}$, we can write $B_i = (B_i \cap E^{\mathsf{c}}) \dot{\bigcup} (B_i \cap E)$ as a disjoint union of sets in $\mathscr{E}$. Using the fact that μ_0 is a pre-measure, we get

$$\mu_0(B_i) = \mu_0(B_i \cap E^{\mathsf{c}}) + \mu_0(B_i \cap E).$$

Thus,

$$\begin{aligned}\mu^*(A) + \epsilon &> \sum_{i=1}^{\infty} \mu_0(B_i \cap E) + \sum_{i=1}^{\infty} \mu_0(B_i \cap E^{\mathsf{c}}) \\ &\geq \mu^*(A \cap E) + \mu^*(A \cap E^{\mathsf{c}}).\end{aligned}$$

Since ϵ was arbitrary we get the result.

□

Now we are ready to prove the extension theorem i.e given an algebra $\mathscr{E}$ on a set X, and a pre-measure μ_0 we can extend the pre-measure to a measure μ on a σ- algebra containing $\mathscr{E}$. Such a construction can be achieved in two step:

(1) First extend the pre-measure μ_0 on $\mathscr{E}$ to an outer measure μ^* on $\mathscr{P}(X)$. See 3.4.2 and 3.4.2.

(2) Then restrict the outer measure μ^* to a σ- algebra $\mathscr{M}$. See 3.4.1.

This is an important result, since we start with simple set functions like measuring length of interval etc. and want to extend this notion to a larger class of sets. The following theorem combines these observations and is called the Carathéodory Extension theorem.

Theorem 3.4.3 (Carathéodory Extension Theorem)**.** Let X be a non-empty set and let $\mathscr{E} \subset \mathscr{P}(X)$ be an algebra, μ_0 a pre-measure on $\mathscr{E}$ and $\sigma(\mathscr{E})$ be the σ- algebra generated by $\mathscr{E}$. Then,

(1) There exists a measure μ on $\sigma(\mathscr{E})$ whose restriction to $\mathscr{E}$ is μ_0.
(2) If ν is another measure on $\sigma(\mathscr{E})$ that extends μ_0, then $\nu(E) \leq \mu(E)$ for any $E \in \sigma(\mathscr{E})$, with equality when $\mu(E) < \infty$.
(3) Additionally, if μ_0 is σ- finite, then μ is the unique extension of μ_0 to a measure on $\sigma(\mathscr{E})$.

Proof. Given non-empty set X, an algebra $\mathscr{E}$ with a pre-measure μ_0,

(1) first extend μ_0 to μ^*, 3.4.2, 3.4.2. Then restrict μ^* to μ on σ- algebra $\mathscr{M}$, 3.4.1. Since $\mathscr{M}$ contains $\mathscr{E}$, it contains $\sigma(\mathscr{M})$. Thus there is a measure μ on $\sigma(\mathscr{E})$ such that,
$$\mu = \mu^*\big|_{\sigma(\mathscr{E})},$$
and
$$\mu\big|_{\mathscr{E}} = \mu^*\big|_{\mathscr{E}} = \mu_0 \,.$$
(2) Note that,
$$\nu\big|_{\mathscr{E}} = \mu_0 = \mu\big|_{\mathscr{E}}.$$
Thus for any $A \in \mathscr{E}$ we have $\nu(A) = \mu_0(A) = \mu(A)$. However, if $\{A_i\}$ is a collection of sets $A_i \in \mathscr{E}$, $\bigcup\limits_{i\in\mathbb{Z}^+} A_i$ may or may not belong in $\mathscr{E}$. In that case, it is not clear that if, $A = \bigcup\limits_{i\in\mathbb{Z}^+} A_i$, then $\nu(A) = \mu(A)$. But for any integer n,
$$\nu(\bigcup_{i=1}^{n} A_i) = \mu(\bigcup_{i=1}^{n} A_i).$$

To this end observe that, $A_1 \subset (A_1 \cup A_2) \subset \cdots \subset \bigcup_{i \in \mathbb{Z}^+} A_i$. Thus we have an increasing sequence of sets $\left((\bigcup_{i=1}^{n} A_i) \right)_n$ and since ν, μ are measures we get,

$$\nu(A) = \lim_{n\to\infty} \nu(\bigcup_{i=1}^{n} A_i) = \lim_{n\to\infty} \mu \left(\bigcup_{i=1}^{n} A_i \right) = \mu\,(A)\,.$$

Here we have used 3.2.1 (4) with $E_i = (\bigcup_{j=1}^{i} A_j)$.

Now, consider any $E \in \sigma\,(\mathscr{E})$ such that a collection $\{A_i\} \subset \mathscr{E}$ and $E \subset \bigcup_{i \in \mathbb{Z}^+} A_i$. Then,

$$\nu(E) \le \nu(\bigcup_{i \in \mathbb{Z}^+} A_i) \le \sum_{i=1}^{\infty} \nu(A_i) = \sum_{i=1}^{\infty} \mu_0(A_i).$$

But this means that,

$$\nu(E) \le \inf \left\{ \sum_{i=1}^{\infty} \mu_0(A_i) \ : \ E \subset \bigcup_{i \in \mathbb{Z}^+} A_i \right\} = \mu\,(E)\,,$$

since $E \in \sigma\,(\mathscr{E})$.

For any $\epsilon > 0$, let $A = \bigcup_{i \in \mathbb{Z}^+} A_i$ such that the collection $\{A_i\} \subset \mathscr{E}$ and $E \subset \bigcup_{i \in \mathbb{Z}^+} A_i$ and, $\sum_{i=1}^{\infty} \mu_0(A_i) < \mu\,(E) + \epsilon$. But $\mu_0(A_i) = \mu\,(A_i)$ and $\mu\,(A) \le \sum_{i=1}^{\infty} \mu\,(A_i)$ and hence, $\mu\,(A) < \mu\,(E) + \epsilon$. Now if $\mu\,(E) < \infty$, since $E \subset A$, $\mu\,(A - E) = \mu\,(A) - \mu\,(E) < \epsilon$. Thus,

$$\begin{aligned}
\mu(E) \le & \mu(A) && \text{because } E \subset A \\
& = \nu(A) && \text{we proved this above} \\
& = \nu(E) + \nu(A \cap E^{\mathsf{c}}) && \text{because } A = E \dot{\bigcup} A \cap E^{\mathsf{c}} \\
& \le \nu(E) + \mu(A \cap E^{\mathsf{c}}) && \text{because } \nu(B) \le \mu(B), \forall B \in \sigma\,(\mathscr{E}) \\
& \le \nu(E) + \epsilon
\end{aligned}$$

Hence, $\mu\,(E) = \nu(E)$ whenever $\mu\,(E) < \infty$.

(3) If μ_0 is σ- finite then $X = \bigcup_{i \in \mathbb{Z}^+} A_i$ with $\mu_0(A_i) < \infty$. We can take A_i to be disjoint. (If not we can contruct sequence of disjoint set F_i whose union is X). Take any $E \in \sigma\,(\mathscr{E})$ such that $E \subset \dot{\bigcup}_{i \in \mathbb{Z}^+} A_i$

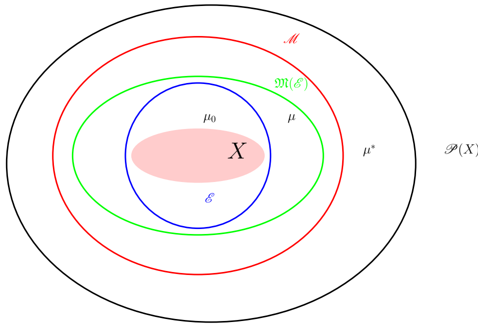

Figure 3.6. Embedding of the various sets. The measures must agree on the *edges*.

and hence $E = \dot{\bigcup}_i (E \cap A_i)$. Hence,

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E \cap A_i) = \sum_{i=1}^{\infty} \nu(E \cap A_i) = \nu(E).$$

Note that the third equality is beacause $\mu(E \cap A_i) \le \mu(A_i) = \mu_0(A_i) < \infty$ and we can use the results from above. Hence $\nu = \mu$.

□

Remark 3.4.1 (Relation between $\mathscr{M}$ and $\sigma(E)$)**.** *This is illustrated in 3.6.*

Example 3.4.2. *Show the usefulness of 3.4.2.*

3.5. Borel Measure on the real line

We have now the machinary to measure all the Borel sets in $\mathbb{R}$. These are the sets that we care about. However we only have a notion of measure for simple sets like intervals, **h-intervals**, etc. Using such elementary notions we want to construct a measure on $\mathfrak{B}_{\mathbb{R}}$. We will begin with a general construction that will yield us the Cumulative Distribution Function (CDF).

To motive the idea, let us say we have a *finite* measure μ on $\mathfrak{B}_{\mathbb{R}}$. Thus we can measure the h-intervals like $(-\infty, x]$. Let $F(x) = \mu((-\infty, x])$. Then we can observe a few facts about F,

(1) (positive real valued) The function F is such that $F : \mathbb{R} \to [0, \infty]$.
(2) (increasing) If $x_1 \le x_2$ then $\mu((-\infty, x_1]) \le \mu((-\infty, x_2])$ from monotonicity of μ. Thus $F(x_1) \le F(x_2)$.
(3) (right continuous) If $x_n \searrow x$ then $(-\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, x_n]$. Thus, $\mu((-\infty, x]) = \lim_{n\to\infty} \mu((-\infty, x_n])$, i.e. $F(x_n) \searrow F(x)$ as $x_n \searrow x$.

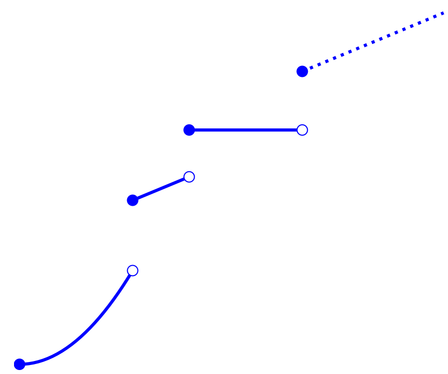

Figure 3.7. A right continuous increasing function

(4) If $y > x$, then $(-\infty, y] = (-\infty, x] \dot{\bigcup} (x, y]$ and hence, $\mu((x, y]) = F(y) - F(x)$.

Now we will turn this around and create a measure on $\mathfrak{B}_{\mathbb{R}}$ using function F with properties as seen above. Such functions are special and we define them below.

Definition 3.5.1 (Monotonic functions)**.** *A real valued function F is* increasing *if $F(x) \leq F(y)$ whenever $x < y$ for all $x, y \in \mathbb{R}$. A real valued function F is* decreasing *if $F(x) \geq F(y)$ whenever $x < y$ for all $x, y \in \mathbb{R}$. A real valued function F is* monotonic *if it is either increasing or decreasing.*

Theorem 3.5.1 (Monotonic functions and one-sided limits)**.** If $F : \mathbb{R} \to \mathbb{R}$ is a monotone function, then F has right and left-hand limits at each point $x \in \mathbb{R}$,

$$F(a^+) = \lim_{x \searrow a} F(x) = \begin{cases} \inf\limits_{x>a} F(x) & \text{If } F \text{ is increasing} \\ \sup\limits_{x<a} F(x) & \text{If } F \text{ is decreasing} \end{cases}$$

$$F(a^-) = \lim_{x \nearrow a} F(x) = \begin{cases} \inf\limits_{x>a} F(x) & \text{If } F \text{ is increasing} \\ \sup\limits_{x<a} F(x) & \text{If } F \text{ is decreasing} \end{cases}$$

Definition 3.5.2 (Right continuous function)**.** *A monotone function*

$$f : \mathbb{R} \to \mathbb{R}$$

is right continuous *if $F(a) = F(a^+)$ for all $a \in \mathbb{R}$, thus,*

$$\lim_{x \searrow a} F(x) = F(a).$$

A right continuous function is show in 3.7. Now we begin our construction of measure on $\mathfrak{B}_{\mathbb{R}}$. We have seen that the collection of h-intervals is an elementary family (3.1.2). A collection of finite disjoint union of h-intervals is, then, an algebra by 3.1.6. If we can define a pre-measure on this algebra, then by 3.4.3, we will have a measure on $\mathfrak{B}_{\mathbb{R}}$ (3.1.4). Thus our first step will be constructing a pre-measure.

Proposition 3.5.1 (Pre-measure on collection of **h-intervals**). *Let $\mathscr{J}_1^{cr}$ be the collection of **h-intervals** on $\mathbb{R}$ and let $\mathscr{A}$ be the algebra of finite disjoint unions of h-intervals. Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. If $(a_j, b_j]$ $(j = 1 \dots n)$ are disjoint h-intervals, define*

$$\mu_0\left(\bigcup_{j=1}^{n}(a_j, b_j]\right) = \sum_{j=1}^{n}(F(b_j) - F(a_j)).$$

Then, μ_0 is a pre-measure.

Proof. First note that if $n = 1$ and if $F(x) = x$, then the μ_0 gives us the *length* of an h-interval. To show that μ_0 is a pre-measure on the algebra $\mathscr{A}$, we need to show that $\mu_0(\emptyset) = 0$ and $\mu_0\left(\dot{\bigcup}_{i=1}^{n} I\right) = \sum_{i=1}^{n}\mu_0(I_i)$. Moreover, if $\dot{\bigcup}_{i=1}^{\infty} I \in \mathscr{A}$, then we need to show that $\mu_0\left(\dot{\bigcup}_{i=1}^{\infty} I_i\right) = \sum_{i=1}^{\infty}\mu_0(I_i)$. Note that $\emptyset = (b, b]$ and thus $\mu_0(\emptyset) = 0$.

We will complete the proof in following steps:

Step 1: Let us check finite additivity. Let $I \in \mathscr{A}$. Then $I = \dot{\bigcup}_{i=1}^{n} I_i$ where $I_i = (a_i, b_i]$. Note that $I_i \in \mathscr{A}$ and hence, $\mu_0(I_i) = F(b_i) - F(a_i)$. Thus,

$$\mu_0(I) = \sum_{i=1}^{n}(F(b_i) - F(a_i)) = \sum_{i=1}^{n}\mu_0(I_i).$$

Hence, μ_0 satisfies finite additivity.

Step 2: We will show that μ_0 is well defined. First let us look at a simple case. Let $I \in \mathscr{A}$, $I = \dot{\bigcup}_{i=1}^{n_I}(a_i, b_i]$. Now assume $I = (a, b]$. If we re-label a_i, b_i, that is $a = a_1 < b_1 = a_2 < b_2 = a_3 \cdots < b$, then $\mu_0(I) = F(b) - F(a)$ from cancellation.

Note, however, such a representation is not unique. We could have a $J \in \mathscr{A}$ such that $J = \dot{\bigcup}_{j=1}^{n_J}(a_j, b_j] = (a, b]$. But $\mu_0(J)$ would also yield $F(b) - F(a)$ by the same argument. See 3.8.

Now, for a general case. If $I = \dot{\bigcup}_{i=1}^{n_I} I_i = \dot{\bigcup}_{j=1}^{n_J} J_j$, then $I_i \subset \dot{\bigcup}_{j=1}^{n_J} J_j$ and thus $I_i = \dot{\bigcup}_{j=1}^{n_J}(I_i \cap J_j)$. But $I_i \in \mathscr{A}$, $I_i \cap J_j$ is an h-interval and hence $\mu_0(I_i) = \sum_{j=1}^{n_J}\mu_0(I_i \cap J_j)$. We have, $\mu_0(I) = \sum_{i=1}^{n_I}\mu_0(I_i)$ and hence, $\mu_0(I) =$

$\sum\limits_{i=1}^{n_I}\sum\limits_{j=1}^{n_J} I_i \cap J_j$. We can repeat this argument by noticing that $J_j \subset \dot{\bigcup}_{i=1}^{n_I} I_i$. Hence, we get

$$\mu_0(I) = \sum_{i=1}^{n_I}\sum_{j=1}^{n_J} I_i \cap J_j = \mu_0(J).$$

Thus μ_0 is well defined.

Step 3: We will show that if $I = \dot{\bigcup}_{i=1}^{\infty} I_i \in \mathscr{A}$, then $\mu_0(I) = \mu_0\left(\dot{\bigcup}_{i=1}^{\infty} I_i\right) \geq \sum\limits_{i=1}^{\infty} \mu_0(I_i)$.

If $I = \dot{\bigcup}_{i=1}^{\infty} I_i \in \mathscr{A}$ then $I = \dot{\bigcup}_{k=1}^{N} (a_k, b_k]$. Since each I_i's and each $(a_k, b_k]$ are disjoint we can partition the sequence finitely, i.e for each $1 \leq k \leq N$ pick n in the sequence such that $I_n \subset (a_k, b_k]$. The set $n_k = \{n \in \mathbb{Z}^+ : I_n \subset (a_k, b_k]\}$ indexs a subsequence of I_i. Hence there is atleast one such n_k, lets call it n_0 such that n_0 is infinite, otherwise we will have only a finite sequence of I_i. Without loss of generality we can disregard other $(a_k, b_k]$ and just consider $I = (a_{n_0}, b_{n_0}] = (a, b]$ (after dropping the n_0).

Note that for $n \in \mathbb{Z}^+$, we have from finite additivity,

$$\mu_0(I) = \sum_{i=1}^{n} \mu_0(I_i) + \mu_0\left(\dot{\bigcup}_{i>n} I_i\right).$$

Since F is increasing the last term is always greater than zero and thus,

$$\mu_0(I) \geq \sum_{i=1}^{n} \mu_0(I_i).$$

This is true for any $n \in \mathbb{Z}^+$ and thus taking $n \to \infty$, we get

$$\mu_0(I) \geq \sum_{i=1}^{\infty} \mu_0(I_i).$$

Step 4: We will show that $\mu_0(I) \leq \sum\limits_{i=1}^{\infty} \mu_0(I_i)$. Pick any $\epsilon > 0$. Note that $I = (a, b]$. Now assume $-\infty < a < b < \infty$. Since F is right continuous at a, there is a $\delta > 0$ such that $F(a+\delta) - F(a) < \epsilon$, but this means that $\mu_0((a, a+\delta]) < \epsilon$. Now, $I = (a, a+\delta] \dot{\bigcup} (a+\delta, b]$ and since $I \in \mathscr{A}$, we get

$$\begin{aligned}\mu_0(I) &= \mu_0((a, a+\delta]) + \mu_0((a+\delta, b]) \\ &< \epsilon + \mu_0((a+\delta, b])\end{aligned}$$

Now that we have the right inequality we need to worry about $\mu_0((a+\delta, b])$. We will use a compactness argument. Note that $[a+\delta, b]$ is closed and bounded subset of $\mathbb{R}$ (from our assumption about a,b) and is thus compact.

Let us use I_i's to create an open cover. However each I_i is an h-interval. We will use the right continuity of F to create open intervals from each I_i. Since F is right continuous at each b_i, there is a $\delta_i > 0$ such that $F(b_i+\delta_i) - F(b_i) < \frac{\epsilon}{2^i}$. Thus, $F(b_i+\delta_i) < F(b_i) + \frac{\epsilon}{2^i}$. Let

$$\mathscr{G} = \left\{(a_i, b_i+\delta_i) \ : \ i \in \mathbb{Z}^+\right\}.$$

be a family of open sets. Then, $[a+\delta, b] \subset \bigcup_{i\in\mathbb{Z}^+} (a_i, b_i+\delta_i)$. Thus $\mathscr{G}$ is an open cover for the compact $[a+\delta, b]$, and thus there exists a finite subcover,

$$\mathscr{G}_N = \{(a_i, b_i+\delta_i) \ : \ 1 \leq i \leq N\}.$$

(We discard any such open set that is contained in a larger open set). After relabeling, we can assume that,

$$b_i + \delta_i \in (a_{i+1}, b_{i+1}+\delta_{i+1}), 1 \leq i \leq N.$$

See 3.9. Thus $F(b_i+\delta_i) \geq F(a_{i+1})$. Note that $b < b_N + \delta_N$ and $a+\delta > a_1$, and thus since F is increasing $F(b) \leq F(b_N+\delta_N)$ and $F(a+\delta) \geq F(a_1)$. Thus,

$$\begin{aligned}
\mu_0\,(I) <& \epsilon + \mu_0\,((a+\delta, b]) \\
&= \epsilon + F(b) - F(a+\delta) \\
&\leq \epsilon + F(b_N+\delta_N) - F(a_1) \\
&= \epsilon + F(b_N+\delta_N) - F(a_N) + \sum_{i=1}^{N-1} (F(a_{i+1}) - F(a_i)) \\
&\leq \epsilon + F(b_N+\delta_N) - F(a_N) + \sum_{i=1}^{N-1} (F(b_i+\delta_i) - F(a_i)) \\
&< \epsilon + \sum_{i=1}^{N} (F(b_i) + \frac{\epsilon}{2^i} - F(a_i)) \\
&= \epsilon + \sum_{i=1}^{N} (\mu_0\,(I_i) + \frac{\epsilon}{2^i}) \\
&< \sum_{i=1}^{\infty} (\mu_0\,(I_i)) + 2\epsilon.
\end{aligned}$$

Since ϵ was arbitrary we get our result. □

Now we apply the framework for constructing measures for the Borel Sets.

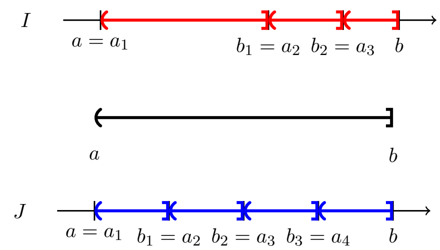

Figure 3.8. Illustration of proof 3.5.1-well defined

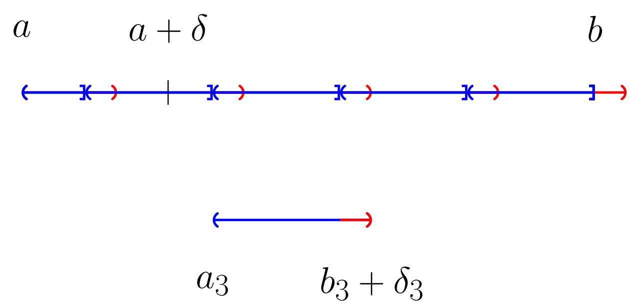

Figure 3.9. Illustration of proof 3.5.1-finite sub-cover

Theorem 3.5.2 (Borel measure on the real line)**.** If $F : \mathbb{R} \to \mathbb{R}$ is any increasing, right continuous function, there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$. If G is another such function, we have $\mu_F = \mu_G$ iff $F - G$ is a constant. Conversely, if μ is a Borel measure on $\mathbb{R}$ that is finite on all bounded Borel sets and we define,

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -\mu((-x,0]) & \text{if } x < 0, \end{cases}$$

then F is increasing and right continuous, and $\mu = \mu_F$.

Proof. The 3.5.1 implies that F induces a pre measure μ_0 on the algebra $\mathscr{A}$ of finite disjoint unions of h intervals. Thus, it can be extended by 3.4.3 to a measure μ_F on σ- algebra generated by $\mathscr{A}$ which is $\mathfrak{B}_{\mathbb{R}}$.

When F, G differ by a constant they give rise to the same pre measure μ_0 on $\mathscr{A}$. Also μ_0 is σ- finite since,

$$\mathbb{R} = \bigcup_{i=-\infty}^{\infty} (i, i+1],$$

hence by 3.4.3 they are equal on $\mathfrak{B}_{\mathbb{R}}$.

As for the converse, we follow the same argument as the observations noted at the beginning of this section. The monotonicity of μ makes F increasing. Since μ is continuous from above and below, we get the right continuity of F. Moreover $\mu(a, b] = F(b) - F(a) = \mu_F$ on $\mathscr{A}$ and thus $\mu = \mu_F$ by 3.4.3.

$\square$

Remark 3.5.1. *From 3.4.1, we know that $\mathfrak{B}_R$ is not necessarily complete. We can complete it to yield the set of all μ^* - measurable sets. This completion* expands *the Borel sets and yields what is called the* Lebesgue *measurable sets. We give a precise definition and few observations in the subsection below.*

The Lebesgue-Stieltjes measure in $\mathbb{R}$.

Definition 3.5.3 (Lebesgue-Stieltjes measure)**.** *If $F : \mathbb{R} \to \mathbb{R}$ is increasing and right continuous, the* complete *measure of μ_F denoted by $\overline{\mu_F}$ is called the Lebesgue-Stieltjes (L-S) measure. We usually drop the overbar and denote it by the same μ_F. The L-S measure associated by the function $F(x) = x$ is called the* Lebesgue *measure and is denoted by $\mu_x = \mu_{\mathscr{L}} = \mu$. We denote its domain by $\mathscr{L}$, and call it the class of **Lebesgue measurable sets**.*

We make some observations about the Lebesgue and Lebesgue-Stieltjes measures. We fix a monotone increasing, right continuous function $F : \mathbb{R} \to \mathbb{R}$, the associated *complete* measure $\mu = \mu_F$, and the domain $\mathscr{M}$ of μ. When $F(x) = x$, we denote $\mathscr{M}$ as $\mathscr{L}$. The measure space is thus the triple $(\mathbb{R}, \mathscr{M}, \mu)$. For any $E \subset \mathscr{M}$, μ is defined as,

$$\mu(E) = \inf\left\{\sum_{i=1}^{\infty} \mu((a_i, b_i]) \ : \ E \subset \bigcup_{i=1}^{\infty} (a_i, b_i]\right\}, \tag{3.1}$$

where $\mu((a, b]) = F(b) - F(a)$.

Theorem 3.5.3 (Equivalent characterizations of L-S measures)**.** If $(\mathbb{R}, \mathscr{M}, \mu)$ is a L-S measurable space then for any $E \in \mathscr{M}$,

- (Measure through open intervals)

$$\mu(E) = \inf\left\{\sum_{i=1}^{\infty} \mu((a_i, b_i)) \ : \ E \subset \bigcup_{i=1}^{\infty} (a_i, b_i)\right\}, \tag{3.2}$$

- (Measure through open sets)

$$\mu(E) = \inf\{\mu(G) \ : \ E \subset G, \text{ G is open}\}, \tag{3.3}$$

- (Measure through compact sets)

$$\mu(E) = \sup\{\mu(K) \ : \ E \supset K, \text{ K is compact}\}. \tag{3.4}$$

Proof. Let us first prove 3.2. Given a collection of open intervals (a_i, b_i), such that $E \subset \bigcup\limits_{i=1}^{\infty} (a_i, b_i)$, we construct a sequence of h-intervals,

$$(a_i, b_i) = \dot{\bigcup}_{k=1}^{\infty} (a_i, b_i - 1/k] = \dot{\bigcup}_{k=1}^{\infty} I_i^k,$$

where, $I_i^k = (a_i, b_i - 1/k]$. Therefore,

$$E \subset \bigcup_{i,k}^{\infty} I_i^k,$$

and hence from 3.1 we get,

$$\mu(E) \leq \sum_{i,k=1}^{\infty} \mu\left(I_i^k\right).$$

From our construction we see, since each $(a_i, b_i - 1/k]$ are disjoint, (to be formal, note that $I_i^1 \subset I_i^1 \dot{\bigcup} I_i^2 \ldots$, thus $\mu\left(\dot{\bigcup\limits_{k=1}^{\infty}} I_i^k\right) = \lim\limits_{n\to\infty} \mu\left(\dot{\bigcup}_{k=1}^{n} I_i^k\right)$),

$$\begin{aligned}
\mu((a_i, b_i)) &= \lim_{n\to\infty} \sum_{k=1}^{n} \mu((a_i, b_i - 1/k]) \\
&= \sum_{k=1}^{\infty} \mu((a_i, b_i - 1/k])
\end{aligned}$$

And thus,

$$\sum_{i,k=1}^{\infty} \mu\left(I_i^k\right) = \sum_{i=1}^{\infty} \mu((a_i, b_i))$$

Hence,

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu((a_i, b_i)).$$

To get the other inequality, fix an $\epsilon > 0$. Hence there is a sequence of $I_i = (a_i, b_i]$ such that $E \subset \bigcup\limits_{i=1}^{\infty} I_i$ and

$$\sum_{i=1}^{\infty} \mu(I_i) < \mu(E) + \epsilon.$$

Since, F is right continuous at every b_i, there is $\delta_i > 0$ such that $F(b + \delta_i) - F(b) < \frac{1}{2^i}$. Now, $(a_i, b_i + \delta_i) \subset (a_i, b_i + \delta_i]$, thus $\mu((a_i, b_i + \delta_i)) \leq \mu((a_i, b_i + \delta_i])$. We have,

$$\begin{aligned}
\mu((a_i, b_i + \delta_i]) &= F(b_i + \delta_i) - F(a_i) \\
&= F(b_i + \delta_i) - F(b_i) + F(b_i) - F(a_i) \\
&< \frac{1}{2^i} + \mu((a_i, b_i]). \\
&= \frac{1}{2^i} + \mu(I_i)
\end{aligned}$$

Hence,

$$\sum_{i=1}^{\infty} \mu((a_i, b_i)) \leq \sum_{i=1}^{\infty} \mu(I_i) < \mu(E) + \epsilon.$$

Since ϵ was arbitrary we get the result.

Now we will prove 3.3. Clearly, if $E \subset G = \bigcup_{i=1}^{\infty}(a_i, b_i)$ then, $\mu(E) \leq \mu(G)$. Fix an $\epsilon > 0$, then from 3.2, there is a collection of open intervals (a_i, b_i) such that $E \subset \bigcup_{i=1}^{\infty}(a_i, b_i)$ and,

$$\sum_{i=1}^{\infty} \mu((a_i, b_i)) < \mu(E) + \epsilon.$$

Put $G = \bigcup_{i=1}^{\infty}(a_i, b_i)$. Then G is open and $E \subset G$. Also, $\mu(G) \leq \sum_{i=1}^{\infty} \mu((a_i, b_i))$. Hence, the result follows.

As for 3.4, note that we are measuring from *inside*. Compact sets in $\mathbb{R}$ are closed and bounded and thus are in $\mathfrak{B}_{\mathbb{R}}$ and thus in $\mathscr{M}$. So for any $K \subset E$, where K is compact, $\mu(K) \leq \mu(E)$ is evident. Thus, we need to show the other direction. We will do it by cases.

CASE I:

E is bounded. If E is also closed then the result is obvious ($K = E$). Thus assume E is not closed. Hence, for any $\epsilon > 0$, we need to find a K compact such that $K \subset E$ and $\mu(K) > \mu(E) - \epsilon$. Since E is not closed $\overline{E} - E$ is not empty. Using equation 3.3 there is an open set G such that $\overline{E} - E \subset G$ and $\mu(G) < \mu\big((\overline{E} - E)\big) + \epsilon$. Let $K = \overline{E} - G$, then K is compact since it is the intersection of two closed sets and is bounded. See 3.10. Also $K \subset E$. From 3.10, it is also evident that $E \cap K^{\mathsf{c}} = E \cap G$, i.e

$$\begin{aligned} E \cap K^{\mathsf{c}} =& E \cap (\overline{E}^{\mathsf{c}} \cup G) \\ =& E \cap G \end{aligned}$$

Now, since $E = K \dot{\bigcup} (E \cap K^{\mathsf{c}}) = K \dot{\bigcup} (E \cap G)$, we have,

$$\begin{aligned} \mu(K) &= \mu(E) - \mu(E \cap G) \\ &= \mu(E) - (\mu(G) - \mu(G - E)) \\ &\geq \mu(E) - \mu(G) + \mu\left(\overline{E} - E\right) \\ &> \mu(E) - \epsilon. \end{aligned}$$

CASE II:

If E is unbounded, then $E_j = E \cap (-j-1, j]$ is bounded. Also $E_j \nearrow E$, hence $\mu(E) = \lim_{j \to \infty} \mu(E_j)$. If $\mu(E) = \infty$, then the result is trivial. Hence, assume $\mu(E) < \infty$. Then for any ϵ, there is a N such that,

$$\mu(E) < \mu(E_N) + \frac{\epsilon}{2}.$$

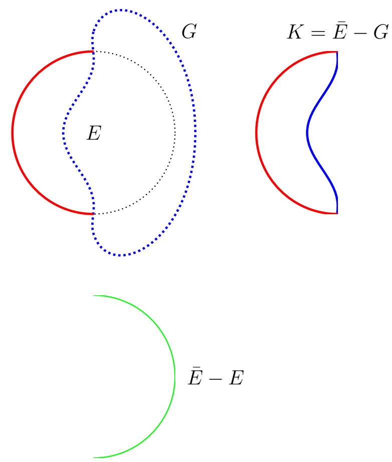

Figure 3.10. Illustration of proof of 3.4-compact measure

Also since E_N is bounded, by the preceding argument, there is a compact set K_N such that $K_N \subset E_N$ and

$$\mu\left(K_N\right) > \mu\left(E_N\right) + \frac{3\epsilon}{2},$$

and thus

$$\mu\left(K_N\right) > \mu\left(E\right) + \epsilon.$$

Thus we have shown that, for any $\epsilon > 0$, there is a compact set $K_N \subset E$, such that

$$\mu\left(K_N\right) > \mu\left(E\right) + \epsilon.$$

$\square$

Note that, the outer measure on $\mathscr{P}(\mathbb{R})$ is defined as,

$$\mu^*\left(A\right) = \inf\left\{\sum_{i=1}^{\infty}\mu_0\left(I_i\right) \;:\; \{I_i\} \subset \mathscr{E} \text{ and } A \subset \bigcup_{i=1}^{\infty} I\right\},$$

where $\mathscr{E}$ is the algebra of finite disjoint union of h-intervals, and $\mu_0\left(I_i\right) = F(b_i) - F(a_i)$, where $I_i = (a_i, b_i]$. The set A may or may not be in $\mathscr{M}$. For it to be measurable, it must satisfy Carathéodory criteria. Whenever $A \in \mathscr{M}$, i.e A is L-S measurable, then we know $\mu^*\left(A\right) = \mu\left(A\right)$. In the theorem below, we give an equivalent criteria for a set $A \subset \mathbb{R}$ to be $L - S$ measurable.

Theorem 3.5.4 (Equivalent criteria for (L-S) measurablity)**.** A set $A \subset \mathbb{R}$ is (L-S) measurable iff for every $\epsilon > 0$ there is an open set $G_\epsilon \supset A$ such that,

$$\mu^*\left(G_\epsilon - A\right) < \epsilon.$$

Proof. Assume $A \subset \mathbb{R}$ is $(L-S)$ measurable. Then $\mu^*(A) = \mu(A)$. Fix an $\epsilon > 0$. From 3.3 there is an open set G such that,

$$\mu(G_\epsilon) < \mu(A) + \epsilon.$$

Also A satisfies Carathéodory criteria,

$$\mu^*(G_\epsilon) = \mu^*(G_\epsilon \cap A) + \mu^*(G_\epsilon \cap A^{\mathsf{c}}).$$

Now, $A \cap G_\epsilon = A$. Assume $\mu^*(A) < \infty$. Then,

$$\mu^*(G_\epsilon \cap A^{\mathsf{c}}) = \mu^*(G_\epsilon) - \mu^*(A) = \mu(G_\epsilon) - \mu(A),$$

since G_ϵ is an open set, hence $G_\epsilon \in \mathscr{M}$ and thus $\mu^*(G_\epsilon) = \mu(G_\epsilon)$. But $\mu(G_\epsilon) - \mu(A) < \epsilon$. Hence, we get the result. If $\mu(A) = \infty$, then pick $A_j = A \cap (-j-1, j]$. Since each A_j is bounded, by the preceding argument there is an open set $G_j \supset A_j$ such that $\mu(G_j - A_j) < \frac{\epsilon}{2^j}$. Let $G_\epsilon = \bigcup_{j=1}^{\infty} G_j$. Now,

$$\begin{aligned}
\mu(G_\epsilon - A) &= \mu\left((\bigcup_{j=1}^{\infty} G_j) - A\right) \\
&= \mu\left(\bigcup_{j=1}^{\infty} (G_j - A)\right) \\
&\leq \mu\left(\sum_{j=1}^{\infty} (G_j - A)\right) \\
&\leq \mu\left(\sum_{j=1}^{\infty} (G_j - A_j)\right) \\
&\leq \epsilon.
\end{aligned}$$

Assume that $A \subset \mathbb{R}$ and for any $\epsilon > 0$ there is an open set G_ϵ such that $\mu^*(G_\epsilon - A) < \epsilon$. We need to show that A is μ^*- measurable. Let $E \subset \mathbb{R}$ be any arbitrary set. The main idea is to show that if A is measurable it will split E w.r.t μ^*. See 3.11. This motivates the statemet, $E - A = (E - G) \dot{\bigcup} (E \cap (G - A))$. Since G is μ^*- measurable we have,

$$\mu^*(E) \geq \mu^*(E \cap G) + \mu^*(E \cap G^{\mathsf{c}}).$$

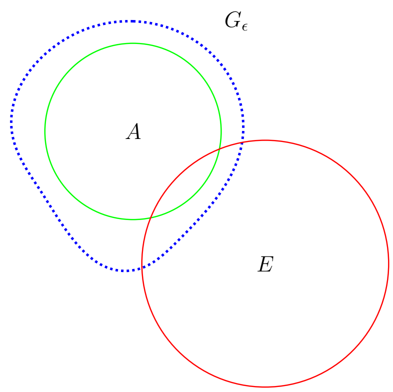

Figure 3.11. Illustration of proof of 3.5.4- part of E not in A is the sum of two parts.

Then from monotonicity of μ^* and noting that $E \cap (G - A) \subset G - A$, $E \cap A \subset E \cap G$, we get

$$\begin{aligned}\mu^*(E \cap A) + \mu^*(E \cap A^{\mathsf{c}}) &= \mu^*(E \cap A) + \mu^*\left((E - G) \dot{\bigcup} (E \cap (G - A))\right) \\ &\leq \mu^*(E \cap A) + \mu^*(E - G) + \mu^*(E \cap (G - A)) \\ &\leq \mu^*(E \cap G) + \mu^*(E - G) + \mu^*(G - A) \\ &< \mu^*(E) + \epsilon.\end{aligned}$$

Hence we get the result. □

The 3.5.4 states that a set is $(L - S)$ measurable if and only if it can be approximated from *outside* by an open set in such a way that the difference has arbitrarily small outer measure. This condition can be adopted as the criteria for a μ^* - measurable set. However, the Carathéodory criteria is very general and is extremely useful for construction of other measures.

The following theorem gives another characterization of $L{-}S$ measurable sets, as ones that can be *squeezed* between open and closed sets.

Theorem 3.5.5 (Squeezing a measurable set by open and closed set)**.** A subset $A \in \mathbb{R}$ is $(L - S)$ measurable if and only if for every $\epsilon > 0$, there is an open set G_ϵ and a closed set F_ϵ such that $G_\epsilon \supset A \supset F_\epsilon$ and,

$$\mu(G_\epsilon - F_\epsilon) < \epsilon.$$

Proof. Assume that for every ϵ there is an open set and closed set G_ϵ, F_ϵ respectively, such that, $G_\epsilon \supset A \supset F_\epsilon$ and

$$\mu(G_\epsilon - F_\epsilon) < \epsilon.$$

From monotonicity of μ^*,

$$\mu^*(G_\epsilon - A) < \mu^*(G_\epsilon - F_\epsilon) = \mu(G_\epsilon - F_\epsilon) < \epsilon,$$

since $G_\epsilon, F_\epsilon \in \mathscr{M}$. Thus, from 3.5.4, A is measurable.

Now, assume $A \subset \mathbb{R}$ is $(L-S)$ measurable. Fix an $\epsilon > 0$. Then there is an open set G_ϵ such that $\mu(G_\epsilon - A) < \epsilon/2$. Since A is $(L-S)$ measurable, A^c is also $(L-S)$ measurable and hence there is an open set $H_\epsilon \supset A^\mathsf{c}$ such that $\mu(H_\epsilon - A^\mathsf{c}) < \epsilon/2$. Let $F_\epsilon = {H_\epsilon}^\mathsf{c}$. Thus $F_\epsilon \subset A$. Note that $A \cap {F_\epsilon}^\mathsf{c} = H_\epsilon \cap A = H_\epsilon - A^\mathsf{c}$. Hence,

$$\begin{aligned}\mu(G_\epsilon - F_\epsilon) &= \mu(A - F_\epsilon) + \mu(G_\epsilon - A)\\ &< \epsilon/2 + \epsilon/2\\ &= \epsilon.\end{aligned}$$

□

Remark 3.5.2. *In the theorem above, if $A \in \mathscr{M}$ and $\mu(A) < \infty$, then we can replace the closed F_ϵ by a compact K_ϵ using 3.4.*

The next theorem states that any Borel set can be approximated up to a set of measure zero by a Borel set.

Theorem 3.5.6 (Borel Approximation)**.** Suppose that $A \subset \mathbb{R}$ is $(L-S)$ measurable. Then there exists a G_δ and F_σ set in $\mathfrak{B}_\mathbb{R}$ such that,

$$G_\delta \supset A \supset F_\sigma, \quad \mu(G_\delta - A) = \mu(F_\sigma - A) = 0.$$

Proof. Since, A is $(L-S)$ measurable, for every $k \in \mathbb{Z}^+$, there is an open set G_k and a closed set F_k such that $G_k \supset A \supset F_k$ and $\mu(G_k - F_k) \leq \frac{1}{k}$. Put $G_\delta = \bigcap_{k=1}^{\infty} G_k$ and $F_\sigma = \bigcup_{k=1}^{\infty} F_k$. Fix an $\epsilon > 0$. For every k,

$$\begin{aligned}G_\delta \cap A^\mathsf{c} &\subset G_k \cap A^\mathsf{c}\\ &\subset G_k \cap F_k^\mathsf{c}\end{aligned}$$

Thus picking a large k such that $\frac{1}{k} < \epsilon$, $\mu(G_\delta - A) \leq \mu(G_k - F_k) < \epsilon$. Since ϵ was arbitrary, the result follows. Similarly,

$$\begin{aligned}A \cap F_\sigma^\mathsf{c} &\subset G_k \cap F_\sigma^\mathsf{c}\\ &\subset G_k \cap F_k^\mathsf{c},\end{aligned}$$

and the result follows. □

3.6. Borel and Lebesgue Measure in $\mathbb{R}^n$

Now we will extend the Borel measure on the real line to $\mathbb{R}^n$. If we view $\mathbb{R}^n$ as $\mathbb{R}\times\mathbb{R}\cdots\times\mathbb{R}$, then we can think of $\mathbb{R}^n$ as the n dimensional product of the Real line. However, we do not have the machinary yet to deal with product measures. We will have to build the Borel measure as we did in 3.5.1. This will give rise to a distribution function in $\mathbb{R}^n$ and the Lebesgue-Stieltjes measure will follow.

The key concept in that construction was that of the distribution function. Thus, given a finite measure on $\mathbb{R}$ we were able to construct an increasing, right continuous function F on $\mathbb{R}$ and vice versa. See 3.5.2. Our goal would be to extend that idea to $\mathbb{R}^n$. Once we have a borel measure on $\mathbb{R}^n$, the complete measure associated with it would give us the Lebesgue-Stieltjes measure on $\mathbb{R}^n$. In $\mathbb{R}$ we looked at an elementary family of h-intervals. In $\mathbb{R}^n$, it will be the family of right semi-closed intervals (rectangles) $\mathscr{I}_n^{cr}$. However, as with $\mathbb{R}$ we need to include $-\infty, \infty$ and these will give us the generalized **h-intervals**.

If $\boldsymbol{a} = (a_1, \ldots, a_n), \boldsymbol{b} = (b_1, \ldots, b_n)$ are points in $\mathbb{R}^n$, let us define $(\boldsymbol{a}, \boldsymbol{b}]$ as the set $\{\boldsymbol{x} \in \mathbb{R}^n \ : \ a_i < x_i \leq b_i, \, 1 \leq i \leq n\}$. Similarly $(\boldsymbol{a}, \infty) = \{\boldsymbol{x} \in \mathbb{R}^n \ : \ x_i > a_i, \, 1 \leq i \leq n\}$ and $(-\infty, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R}^n \ : \ x_i \leq b_i, \, 1 \leq i \leq n\}$. Let $\mathscr{E} \subset \mathbb{R}^n$ be the collection of all generalized h-intervals as described above. Then $\mathscr{E}$ is an elementary family and by 3.1.6 the family $\mathscr{A}$ of finite disjoint union of generalized h-intervals in $\mathscr{E}$ is an algebra. Moreover, $\sigma(\mathscr{A}) = \mathfrak{B}_{\mathbb{R}^n}$. Thus, we have to define a pre-measure μ_0 on $\mathscr{A}$ and by 3.4.3 we will have a unique measure on $\mathfrak{B}_{\mathbb{R}^n}$. We proceed as in 3.5.1, but we need to define what is means for a function F in $\mathbb{R}^n$ to be increasing and right continuous. The notion of a distribution function is more complicated in $\mathbb{R}^n$. As with our motivation in the previous section, let us assume we have a finite measure (for example the Lebesgue measure on a cube) in $\mathbb{R}^n$. In $\mathbb{R}$, we defined $F(x) = \mu((-\infty, x])$. Let us extend the same idea and define $F : \mathbb{R}^n \to \mathbb{R}$ as,

$$F(x_1, x_2, \ldots, x_n) = \mu\left(\{\boldsymbol{\omega} \in \mathbb{R}^n \ : \ \omega_i \leq x_i\}\right).$$

In $\mathbb{R}$, this lead to $\mu((a, b]) = F(b) - F(a)$. However, this wont be true in $\mathbb{R}^n$. We need some more ideas to make a jump from the real line to $\mathbb{R}^n$. One is the notion of order. In $\mathbb{R}^n$ there is no total order, however the following notion will suffice,

Definition 3.6.1 (Order in $\mathbb{R}^n$)**.** *Let $\boldsymbol{a}, \boldsymbol{b}$ be points in $\mathbb{R}^n$. We say that $\boldsymbol{a} \leq \boldsymbol{b}$ iff $a_i \leq b_i$ for all $1 \leq i \leq n$.*

To generalize $F(b) - F(a)$, we need to define the *difference* operator of a function F from $\mathbb{R}^n$ to $\mathbb{R}$.

Definition 3.6.2 (Difference Operator)**.** *Let $G : \mathbb{R}^n \to \mathbb{R}$. We define the difference of G at the ith cordinate evaluated at a_i, b_i as,*

$$\Delta_{b_i a_i} G = G(x_1, x_2, \ldots, x_{i-1}, b_i, \ldots, x_n) - G(x_1, x_2, \ldots, x_{i-1}, a_i, \ldots, x_n).$$

Definition 3.6.3 (Increasing Function in $\mathbb{R}^n$)**.** *Let $F : \mathbb{R}^n \to R$ and for $\boldsymbol{a} \leq \boldsymbol{b}$, let us denote by $F((\boldsymbol{a}, \boldsymbol{b}])$ as,*

$$F((\boldsymbol{a}, \boldsymbol{b}]) = \Delta_{b_1 a_1} \ldots \Delta_{b_n a_n} F.$$

The function F is said to be increasing iff $F((\boldsymbol{a}, \boldsymbol{b}]) \geq 0$ whenever $\boldsymbol{a} \leq \boldsymbol{b}$.

The definition may seem peculiar but it is motivated by the following theorem,

Theorem 3.6.1 (Measure and Distribution function in $\mathbb{R}^n$)**.** Let μ be a finite measure on $\mathfrak{B}_{\mathbb{R}^n}$ and define,

$$F(\boldsymbol{x}) = \mu\left((-\infty, \boldsymbol{x}]\right),$$

Then, if $\boldsymbol{a} \leq \boldsymbol{b}$,

$$\mu\left((\boldsymbol{a}, \boldsymbol{b}]\right) = F((\boldsymbol{a}, \boldsymbol{b}]).$$

Proof. We prove it for $n = 3$ to make the notation simpler.

$$\begin{aligned}
\Delta_{b_3 a_3} F &= F(x_1, x_2, b_3) - F(x_1, x_2, a_3) \\
&= \mu\left(\{\boldsymbol{\omega} \;:\; \omega_1 \leq x_1,\, \omega_2 \leq x_2,\, \omega_3 \leq b_3\}\right) \\
&\quad - \mu\left(\{\boldsymbol{\omega} \;:\; \omega_1 \leq x_1,\, \omega_2 \leq x_2,\, \omega_3 \leq a_3\}\right) \\
&= \mu\left(\{\boldsymbol{\omega} \;:\; \omega_1 \leq x_1,\, \omega_2 \leq x_2,\, a_3 < \omega_3 \leq b_3\}\right)
\end{aligned}$$

Thus,

$$\begin{aligned}
\Delta_{b_2 a_2} \Delta_{b_3 a_3} F &= \Delta_{b_2 a_2}(F(x_1, x_2, b_3) - F(x_1, x_2, a_3)) \\
&= F(x_1, b_2, b_3) - F(x_1, a_2, b_3) \\
&\quad - F(x_1, b_2, a_3) + F(x_1, a_2, a_3) \\
&= \mu\left(\{\boldsymbol{\omega} \;:\; \omega_1 \leq x_1,\, a_2 < \omega_2 \leq b_2,\, a_3 < \omega_3 \leq b_3\}\right)
\end{aligned}$$

and thus we get the result from another application of the difference operator. □

For defining a right continuous function we need to define a *right* limit. We say that a sequence of points $(\boldsymbol{x}^{(n)})$ in $\mathbb{R}^n$ coverges to a point $\boldsymbol{x}$ from the *right* if for each co-ordinate $x_i^{(n)} \searrow x$. We say that $(\boldsymbol{x}^{(n)})$ *decreases* to $\boldsymbol{x}$ and denote it by $(\boldsymbol{x}^{(n)}) \searrow \boldsymbol{x}$.

Definition 3.6.4 (Right continuous function in $\mathbb{R}^n$)**.** *A function, $F : \mathbb{R}^n \to R$ is right continuous if for a sequence of points $\boldsymbol{x}^{(n)} \searrow \boldsymbol{x}$; $F(\boldsymbol{x}^{\boldsymbol{n}}) \searrow F(\boldsymbol{x})$.*

Definition 3.6.5 (Distribution function in $\mathbb{R}^n$)**.** *A distribution function F is an increasing, right continuous function $F\ :\mathbb{R}^n \to \mathbb{R}$.*

Theorem 3.6.2 (Borel measure in $\mathbb{R}^n$)**.** Let F be a distribution function on $\mathbb{R}^n$ and set,

$$\mu_F = F((\boldsymbol{a}, \boldsymbol{b}]), \quad \boldsymbol{a} \leq \boldsymbol{b}.$$

Then μ_F is a measure on $\mathfrak{B}_{\mathbb{R}^n}$ and is called the borel measure on $\mathbb{R}^n$ induced by F.

Proof. The proof follows the same idea as in 3.5.1 and 3.5.3. □

Definition 3.6.6 (Lebesgue-Stieltjes (L-S) measure on $\mathbb{R}^n$)**.** *Let F be a distribution function in $\mathbb{R}^n$ and μ_F the corresponding borel measure on $\mathbb{R}^n$. The completion of μ_F is called the Lebesgue-Stieltjes measure on $\mathbb{R}^n$ and we also denote is by μ_F.*

Example 3.6.1. *Let $F_1, F_2, \ldots, F_n$ be n distribution function on $\mathbb{R}$, and define,*

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1) \times F_2(x_2) \times \ldots F_n(x_n).$$

F is a distribution function on $\mathbb{R}^n$ with,

$$F((\boldsymbol{a}, \boldsymbol{b}]) = \prod_{i=1}^{n} (F_i(b_i) - F_i(a_i)).$$

When $F_i = x$ for all i, $F((\boldsymbol{a}, \boldsymbol{b}]) = \prod_{i=1}^{n} (b_i - a_i)$ and the corresponding $L-S$ measure is the Lebesgue measure on $\mathbb{R}^n$.

For a given distribution function F, we denote by $(\mathbb{R}^n, \mathscr{M}, \mu)$ as Lebesgue-Stieltjes $(L-S)$ measure space. In particular the space of Lebesgue measure is denoted by $(\mathbb{R}^n, \mathscr{L}, \mu)$. Thus by definition a set $E \subset \mathbb{R}^n$ is $L-S$ measurable if,

$$\mu\,(E) = \inf \left\{ \sum_{i=1}^{\infty} F((\boldsymbol{a}, \boldsymbol{b}]_i) \ : \ E \subset \bigcup_{i=1}^{\infty} (\boldsymbol{a}, \boldsymbol{b}]_i \right\},$$

which is equivalent to

$$\mu\,(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu\,(R_i) \ : \ E \subset \bigcup_{i=1}^{\infty} R_i \right\},$$

where R_i are a sequence of closed rectangle with sides parallel to co-ordinate axes.

It is clear that all equivalent criteria for $(L-S)$ measurability we addressed for $\mathbb{R}$ holds for $\mathbb{R}^n$. Thus 3.5.3 holds in $\mathbb{R}^n$, in particular

$$\mu\,(E) = \inf \left\{ \mu\,(G) \ : \ E \subset G, \quad \text{G is open} \right\}, \tag{3.5}$$

and

$$\mu(E) = \sup\{\mu(K) \ : \ E \supset K, \quad \text{K is compact}\}. \tag{3.6}$$

To prove 3.5, note that if $E \subset G$ then $\mu^*(E) \le \mu^*(G)$. Since both E, G are $(L-S)$ measurable their outer measure is just the $(L-S)$ measure μ. Assume $\mu(E)$ is finite. To prove the other direction, for any ϵ there exist a sequence of rectangles $\{R_i\} \subset \mathbb{R}^n$ such that $\sum_{i=1}^{\infty} \mu(R_i) < \mu(E) + \epsilon$. Since $\mu(E)$ is finite each rectangle R_i is bounded and there is an open rectangle $S_i{}^\circ$ such that $R_i \subset S_i{}^\circ$ and $\mu(S_i) < \mu(R_i) + \frac{\epsilon}{2^i}$. Since $S_i{}^\circ$ is an open set, let $G = \bigcup_{i=1}^{\infty} S_i{}^\circ$. Thus we get our result. 3.6 follows exactly as in 3.5.4 (3) with minor modification in the case that $\mu(E)$ is infinite.

Similarly as in 3.5.4, 3.5.5 and 3.5.6,

Theorem 3.6.3 (($L-S$) Measurability in $\mathbb{R}^n$)**.** A set $E \subset \mathbb{R}^n \in \mathscr{M}$ is $(L-S)$ measurable if and only if for every $\epsilon > 0$,

(1) there is an open set $G_\epsilon \subset \mathbb{R}^n$ such that $G_\epsilon \supset E$ and,
$$\mu^*(G_\epsilon - E) < \epsilon,$$
(2) there is an open set G_ϵ and a closed set F_ϵ, such that $G_\epsilon \supset E \supset F_\epsilon$ and,
$$\mu(G_\epsilon - F_\epsilon) < \epsilon,$$
(3) E differs from a G_δ and a F_σ set by a set of measure zero.

Uniqueness of measures and properties of Lebesgue measures. We will use the $\pi - \lambda$ theorem to discuss the uniqueness of measures and use that to show some important properties of Lebesgue measure.

Theorem 3.6.4. Assume that $(X, \mathscr{M})$ is a measurable space and that $\mathscr{M} = \sigma(\mathscr{E})$, where $\mathscr{E}$ is a $\pi-$ class. Suppose that μ_1, μ_2 are two measures on $\sigma(\mathscr{E})$ such that they are sigma-finite on $\mathscr{E}$. If μ_1, μ_2 agree on $\mathscr{E}$, then they agree on $\sigma(\mathscr{E})$.

Proof. Since measures are sigma-finite on $\mathscr{E}$, there is a sequence of subsets (X_n) such that $X_n \nearrow X$ and $\mu_1(X_n) = \mu_2(X_n) < \infty$ for each n. For any $A \in \sigma(\mathscr{E})$, we observe that,

$$A \subset X = \bigcup_{n=1}^{\infty} X_n,$$

and hence $A = \bigcup_{n=1}^{\infty} (A \cap X_n)$. Thus, to show that $\mu_1(A) = \mu_2(A)$ for any $A \in \sigma(\mathscr{E})$, we need to show that $\mu_1(A \cap X_n) = \mu_2(A \cap X_n)$ for each n;

because if that holds, we get

$$\mu_1(A) = \mu_1(\bigcup_{n=1}^{\infty}(A \cap X_n)) = \lim_{n\to\infty} \mu_1(A \cap X_n) = \lim_{n\to\infty} \mu_2(A \cap X_n) = \mu_2(A),$$

since $(A \cap X_n) \nearrow (A \cap X)$. Thus, let $\mathscr{F}_j$ be defined as the class of all those sets A in $\sigma(\mathscr{E})$ such that:

$$\mathscr{F}_j = \{A \in \sigma(\mathscr{E}) \ : \ \mu_1(A \cap X_n) = \mu_2(A \cap X_n)\}.$$

Since $\mathscr{E}$ is a $\pi-$ class we see that for each j, $\mathscr{E} \subset \mathscr{F}_j$. By construction, $\mathscr{F}_j \subset \sigma(\mathscr{E})$. If we show that $\mathscr{F}_j$ is a $\lambda-$ class, we are done; since then $\mathscr{F}_j \supset \sigma(\mathscr{E})$ by the $\pi - \lambda$ theorem and hence we will get $\mathscr{F}_j = \sigma(\mathscr{E})$ for each j.

It is easy to show that $\mathscr{F}_j$ is a $\lambda-$ class.

(1) Clearly $X \in \mathscr{F}_j$.

(2) Let $A \in \mathscr{F}_j$. Note that

$$\mu_1(X_j) = \mu_1(X_j \cap A) + \mu_1(X_j \cap A^{\mathsf{c}})$$

Since $(X_j \cap A) \subset X_j$, $\mu_1(X_j \cap A)$ is finite. Now,

$$\begin{aligned}\mu_1(A^{\mathsf{c}} \cap X_j) &= \mu_1(X_j) - \mu_1(X_j \cap A)\\ &= \mu_2(X_j) - \mu_2(X_j \cap A)\\ &= \mu_2(A^{\mathsf{c}} \cap X_j).\end{aligned}$$

(3) Let (A_n) be a sequence of pairwise disjoint sets in $\mathscr{F}_j$

$$\begin{aligned}\mu_1(X_j \cap \left(\bigcup_{n=1}^{\infty} A_n\right)) &= \mu_1(\bigcup_{n=1}^{\infty}(X_j \cap A_n))\\ &= \sum_{n=1}^{\infty} \mu_1(X_j \cap A_n)\\ &= \sum_{n=1}^{\infty} \mu_2(X_j \cap A_n)\\ &= \mu_2(\bigcup_{n=1}^{\infty}(X_j \cap A_n))\\ &= \mu_2(X_j \cap \left(\bigcup_{n=1}^{\infty} A_n\right)).\end{aligned}$$

Hence, $\bigcup_{n=1}^{\infty} A_n$ is in $\mathscr{F}_j$.

Thus, $\mathscr{F}_j$ is a $\lambda-$ class for each j. Hence, we get the result. $\square$

Chapter 4

Elements of measure theory: measurable functions

In this chapter we will look at mappings from one measurable space to another. This chapter, along with the preceding one, contains the basic elements of measure theory. This will then enable us to study calculus on a general setting.

4.1. Measurable functions

In topology, a continuous function is one for which the inverse image of open sets are open sets. In similar vein, we will define measurable functions between two measure spaces. The definition is motivated by the construction used by Lebesgue to extend Riemann integration. Integration will be dealt in the next chapter.

Definition 4.1.1 (Measurable functions)**.** *Let $(X, \mathscr{M})$, $(Y, \mathscr{N})$ be two measure spaces. A function,*

$$f \ : X \to Y$$

is $(\mathscr{M}, \mathscr{N})$ − measurable if,

$$f^{-1}(A) \in \mathscr{M} \quad \text{for every}\, A \in \mathscr{N}.$$

Consider the collection $f^{-1}(\mathscr{N}) = \left\{ f^{-1}(A) \ : \ A \in \mathscr{N} \right\}$. This is a sigma algebra by 3.1.4. We call this the induced sigma algebra on X. Thus, we see that if f is $(\mathscr{M}, \mathscr{N})$ − measurable then $f^{-1}(\mathscr{N}) \subset \mathscr{M}$. Note that the inverse image in general is a function $f^{-1} : \mathscr{P}(Y) \to \mathscr{P}(X)$. The definition above states that, if f is measurable then the restriction of the inverse image, $f^{-1}\big|_{\mathscr{N}}$, is a function that maps $\mathscr{N}$ to $\mathscr{M}$.

Since, measurable sets are generally *huge*, the above condition is not very useful in figuring which functions are measurable. The following theorem is hence very useful.

Theorem 4.1.1 (Criteria for measurable functions)**.** Let $(X, \mathscr{M})$, $(Y, \mathscr{N})$ be two measure spaces and let $\mathscr{N} = \sigma(\mathscr{E})$. A function,

$$f : X \to Y$$

is $(\mathscr{M}, \mathscr{N})$ – measurable if and only if,

$$f^{-1}(A) \in \mathscr{M} \quad \text{for every } A \in \mathscr{E}.$$

Proof. *only if* part is immediate since, if f is measurable, the statement is true for all $A \in \mathscr{N}$ and hence certainly true for $A \in \mathscr{E}$.

For the *if* part, consider the set

$$F = \left\{A \subset Y \ : \ f^{-1}(A)\right\} \in \mathscr{M}.$$

This set is not empty by our assumption i.e $\mathscr{F} \supset \mathscr{E}$. Easy to see that it is a σ- algebra. See 3.1.1. Hence, $F \supset \sigma(\mathscr{E}) = \mathscr{N}$. □

Corollary 4.1.1.1. *If X, Y are metric spaces (or topological spaces), every continuous function $f : X \to Y$ is $(\mathfrak{B}_X, \mathfrak{B}_Y)$ – measurable.*

Proof. Let $\mathscr{G}_Y$,$\mathscr{G}_X$ be the class of open sets in Y, X respectively. We know that $\mathfrak{B}_Y = \sigma(\mathscr{G}_Y)$. Since f is continuous, for any open set $G \in \mathscr{G}_Y$, $f^{-1}(G)$ is open in $\mathscr{G}_X \subset \mathfrak{B}_X$. Hence, by 4.1.1 f is $(\mathfrak{B}_X, \mathfrak{B}_Y)$ – measurable. □

Definition 4.1.2 (Real or Complex valued measurable functions)**.** *If $(X, \mathscr{M})$ is a measurable space, a real or complexed valued function is called $\mathscr{M}$ measurable (or just measurable if $\mathscr{M}$ is understood from the context) if it is $(\mathscr{M}, \mathfrak{B}_{\mathbb{R}})$ – measurable or $(\mathscr{M}, \mathfrak{B}_{\mathbb{C}})$ – measurable respectively. Thus the co-domain σ- algebra is always the Borel sigma algebra for real or complex valued functions.*

Definition 4.1.3 (Lebesgue measurable functions)**.** *A real (or complex) valued function,*

$$f : \mathbb{R}^n \to \mathbb{R}\,(or\, f : \mathbb{R}^n \to \mathbb{C}),$$

is Lebesgue measurable if it $(\mathscr{L}, \mathfrak{B}_R)$ – measurable (or $(\mathscr{L}, \mathfrak{B}_C)$ – measurable resp.).

We often want to discuss measurability relative to a subset of a measure space.

Definition 4.1.4 (Restriction of measurable functions)**.** *Let $(X, \mathscr{M})$ and $(Y, \mathscr{N})$ be measurable spaces and $E \in \mathscr{M}$ be any measurable set and f be a function defined on X given by $f : X \to Y$. We say that f is measurable on E if $f^{-1}(A) \cap E \in \mathscr{M}$ for any $A \in \mathscr{N}$. We say that $f\big|_E$ is $(\mathscr{M}_E, N)-$measurable, where $\mathscr{M}_E$ is the restricted sigma algebra on E as described in 3.1.1 (5).*

Proposition 4.1.1. *Let $(X, \mathscr{M})$ be a measurable space and let $f : X \to R$ be a real valued function. Then, the following are equivalent:*

(1) *f is $\mathscr{M}$ measurable.*
(2) *$f^{-1}((a, \infty)) \in \mathscr{M}$ for all $a \in \mathbb{R}$.*
(3) *$f^{-1}([a, \infty)) \in \mathscr{M}$ for all $a \in \mathbb{R}$.*
(4) *$f^{-1}((-\infty, a)) \in \mathscr{M}$ for all $a \in \mathbb{R}$.*
(5) *$f^{-1}((-\infty, a]) \in \mathscr{M}$ for all $a \in \mathbb{R}$.*

Proof. We will show (1) $\iff$ (2) using 4.1.1 and 3.1.5.1. The rest are similar. Consider the family of open rays $\mathscr{E}_5$ as described in 3.1.5.1. Since $\sigma(\mathscr{E}_5) = \mathfrak{B}_{\mathbb{R}}$, by 4.1.1 we get the result. □

Remark 4.1.1. *Sometimes, it will be convenient to consider functions with values in the extended reals $\overline{\mathbb{R}} = [-\infty, \infty]$. In that case we can define the Borel sets in $\overline{\mathbb{R}}$ by,*

$$\mathfrak{B}_{\overline{\mathbb{R}}} = \left\{E \subset \overline{\mathbb{R}} \ : \ E \cap \mathbb{R} \in \mathfrak{B}_{\mathbb{R}}\right\}.$$

Then,

$$\mathfrak{B}_{\overline{\mathbb{R}}} = \sigma(\{(a, \infty]\}) = \sigma(\{[-\infty, a)\}).$$

Thus in the proposition above, a function $f : X \to \overline{\mathbb{R}}$ is measurable if we include $-\infty, \infty$, where it is understood that we take $\mathfrak{B}_{\overline{\mathbb{R}}}$ as the co-domain sigma algebra.

For functions that take values in the extended reals, note that the set $f^{-1}([-\infty, a)) = \{x \in X \ : \ f(x) < a\}$. It is a common practice to denote this set as $\{f < a\}$. Thus from 4.1.1 and 4.1.1, a real valued function is measurable iff $\{f < a\}$, $\{f \leq a\}$, $\{f > a\}$ and $\{f \geq a\}$ are measurable. This notation is used in probability and is also preferrable if we do not want to show explicitly a real valued function or a function that takes values on the extended reals.

Example 4.1.1. *We give a few examples of real valued measurable functions.*

(1) *Consider the counting measure on* $\mathbb{R}$,

$$\mu(E) = \begin{cases} \text{number of elements in } E & \text{if } E \text{ is finite} \\ \infty & \text{if } E \text{ is infinite} \end{cases}$$

Then each set is measurable and hence any function is measurable.

(2) *Define an outer measure* μ^* *on* $\mathbb{R}$ *by*

$$\mu^*(E) = \begin{cases} 0 & \text{if } E = \emptyset \\ 1 & \text{otherwise} \end{cases}$$

Then only $\emptyset$ *and* $\mathbb{R}$ *are measurable. Indeed for any* $E \subset \mathbb{R}$ *where* $E \neq \mathbb{R}, E \neq \emptyset$ *the Carathéodory criteria,*

$$\mu^*(\mathbb{R}) \geq \mu^*(\mathbb{R} \cap E) + \mu^*(\mathbb{R} \cap E^{\mathsf{c}})$$

is NOT satisfied. Consider the set $f(x) = x$. *Since* $\{f > 0\}$ *is not measurable,* f *is not measurable.*

(3) *Consider a measurable space* $(X, \mathscr{M})$. *Let* $f : X \to \mathbb{R}$ *be the function given by* $f(x) = k > 0$ *whenever* $x \in E$ *for some* $E \subset X$, *else* $f(x) = 0$. *Hence the range of* f *is just* $\{0, k\}$ *Then,*

$$\{f > c\} = \begin{cases} \emptyset & \text{if } c \geq k \\ E & \text{if } 0 \leq c < k \\ X & \text{if } c < 0 \end{cases}$$

Hence, f *is measurable iff* E *is measurable.*

(4) *Consider the Lebesgue measure space on* $\mathbb{R}$, *i.e the measurable space* $(\mathbb{R}, \mathscr{L})$ *and the following function* f

$$f(x) = \begin{cases} x^2, & x \leq 1 \\ 2, & x = 1 \\ -2 - x, & x > 1 \end{cases}$$

Then,

$$\{f > c\} = \begin{cases} (-\infty, -2 - c) & c \leq -3 \\ (-\infty, 1] & -3 < c < 0 \\ (-\infty, -\sqrt{c}) \cup (\sqrt{c}, 1] & 0 \leq c < 1 \\ (-\infty, -\sqrt{c}) \cup \{1\} & 1 \leq c < 2 \\ (-\infty, -\sqrt{c}) & c \geq 2. \end{cases}$$

Hence, f *is measurable.*

Appropriate measurable functions agree with compostion. This is proved in the following Proposition.

$$
\begin{tikzcd}
 & (Z,\mathscr{O}) \\
(Y,\mathscr{N}) \arrow[ur, "g"] & \\
(X,\mathscr{M}) \arrow[u, "f"] \arrow[uur, dotted, "g\circ f"'] &
\end{tikzcd}
$$

Figure 4.1. Composition of measurable functions

Proposition 4.1.2 (Composing measurable functions)**.** *If* $(X,\mathscr{M})$*,* $(Y,\mathscr{N})$ *and* $(Z,\mathscr{O})$ *are measurable spaces and* $f\ :X\to Y$ *is* $(\mathscr{M},\mathscr{N})-$ *measurable,* $g\ :Y\to Z$ *is* $(\mathscr{N},\mathscr{O})-$*measurable, then the compositon map* $g\circ f\ :X\to Z$ *is* $(\mathscr{M},\mathscr{O})-$ *measurable.*

Proof. Consider an arbitrary set $C\in\mathscr{O}$. Then $B=g^{-1}(C)\in\mathscr{N}$ and so $f^{-1}(B)\in\mathscr{M}$. Thus $g\circ f^{-1}(C)=f^{-1}\left(g^{-1}(C)\right)\in\mathscr{M}$. □

Remark 4.1.2. *Note, 4.1.2 does not imply that two Lebesgue measurable functions are measurable. If* f *is* $(\mathscr{L},\mathfrak{B}_{\mathbb{R}})-$ *measurable and* g *is* $(\mathscr{L},\mathfrak{B}_{\mathbb{R}})-$ *measurable, then the theorem above doesn't gurantee the Lebesgue measurability of* $g\circ f$ *or* $f\circ g$*, since there is no guarantee that for a set* $C\in\mathfrak{B}_{\mathbb{R}}$*,* $g^{-1}(C)\in\mathscr{L}$*. If* g *is Borel measurable then the issue is resolved. See 4.1.*

A measurable function also induces a measure in the co-domain which in some regards is the *natural* measure. To make this idea concrete, we provide the next proposition.

Proposition 4.1.3 (Induced measure)**.** *Let* f *be* $(\mathscr{M},\mathscr{N})-$ *measurable function that maps from a measure space* $(X,\mathscr{M},\mu)$ *in to a measure space* $(Y,\mathscr{N})$*. For any* $B\in\mathscr{N}$*, set*

$$\mu_f(B)=\mu(f^{-1}(B)).$$

Then $(Y,\mathscr{N},\mu_f)$ *is a measure space.*

$$\begin{tikzcd}
\mathscr{N} \arrow[r, "\mu_f = \mu \circ f^{-1}"] \arrow[d, "f^{-1}"] & {[0,\infty]} \\
\mathscr{M} \arrow[ur, "\mu"'] &
\end{tikzcd}$$

Figure 4.2. Induced measure. Since f is measurable, the inverse image when restricted to $\mathscr{N}$ is a function that maps to $\mathscr{M}$. It is also called the **pull back** measure.

Proof. Since $f^{-1}(\emptyset) = \emptyset$, $\mu_f(\emptyset) = 0$. Let $\{B_i\} \subset \mathscr{N}$ be a sequence of pairwise disjoint sets. Then,

$$\begin{aligned}
\mu_f(\dot{\bigcup}_{i=1}^{\infty} B_i) &= \mu\left(f^{-1}\left(\dot{\bigcup}_{i=1}^{\infty} B_i\right)\right), \\
&= \mu\left(\dot{\bigcup}_{i=1}^{\infty} f^{-1}(B_i)\right), \\
&= \sum_{i=1}^{\infty} \mu\left(f^{-1}(B_i)\right), \\
&= \sum_{i=1}^{\infty} \mu_f(B_i).
\end{aligned}$$

Hence, $(Y, \mathscr{N}, \mu_f)$ is a measure space. □

Even though we have included complex valued functions, we haven't defined yet the borel sigma algebra $\mathfrak{B}_{\mathbb{C}}$. Note that $\mathfrak{B}_{\mathbb{C}} = \mathfrak{B}_{\mathbb{R}^2}$. Since we deal with a complex valued function by dealing with its real and imaginary parts, it will be useful to describe a product sigma algebra.

Remark 4.1.3. *If we have a set $(X, \mathscr{M})$ and a collection of measurable spaces $\{(Y_\alpha, \mathscr{N}_\alpha)\}_{\alpha \in A}$, what is the smallest σ- algebra $\mathscr{M}$ on X with respect to which each map $f_\alpha : X \to Y_\alpha$ is $(\mathscr{M}, \mathscr{N}_\alpha)$ − measurable? Note that ${f_\alpha}^{-1}(\mathscr{N}_\alpha)$ is a sigma algebra on X. Let*

$$\mathscr{E} = \bigcup_{\alpha \in A} {f_\alpha}^{-1}(\mathscr{N}_\alpha).$$

The unions of σ- algebra is not necessarily a σ- algebra. The smallest sigma algebra containing a family is the one generated by it. Hence, $\mathscr{M} = \sigma(\mathscr{E})$ is the smallest sigma algebra on X on which all f_α are measurable. In *particular, if* $X = \prod_{\alpha \in A} Y_\alpha$*, then the co-ordinate maps* $\pi_\alpha : X \to Y_\alpha$ *will be* $(\mathscr{M}, \mathscr{N}_\alpha)-$*measurable for suitably defined* $\mathscr{N}_\alpha$*. This leads us to the definition of a product* σ*- algebra.*

Definition 4.1.5 (Product sigma algebra)**.** *Let* $\{X_\alpha : \alpha \in A\}$ *be an indexed collection of non-empty sets. Set,*

$$X = \prod_{\alpha \in A} X_\alpha, \quad \text{and} \quad \pi_\alpha : X \to X_\alpha,$$

as the co-ordinate map. If $\mathscr{M}_\alpha$ *is a* σ*- algebra on* X_α *for each* $\alpha \in A$*, then we define the product sigma algebra on* X *as the* σ*- algebra generated by*

$$\mathscr{E} = \left\{ {\pi_\alpha}^{-1}(E_\alpha) \; : \; E_\alpha \in \mathscr{M}_\alpha, \alpha \in A \right\}.$$

We denote this σ*- algebra by* $\bigotimes_{\alpha \in A} \mathscr{M}_\alpha$*.*

Again, as with measurable functions we want to only check for sets in the family that generates $\mathscr{M}_\alpha$. This leads to the next proposition.

Proposition 4.1.4. *Suppose that* $\mathscr{M}_\alpha$ *is generated by* $\mathscr{E}_\alpha$*. Then* $\bigotimes_{\alpha \in A} \mathscr{M}_\alpha$ *is generated by,*

$$\mathscr{F}_1 = \left\{ {\pi_\alpha}^{-1}(E_\alpha) \; : \; E_\alpha \in \mathscr{E}_\alpha \, \alpha \in A \right\}.$$

Proof. Clearly $\mathscr{F}_1 \subset \mathscr{E}$, hence $\sigma(\mathscr{F}_1) \subset \sigma(\mathscr{E})$. We will use 3.1.3 to prove the other inclusion. Consider the set,

$$\left\{ E \subset X_\alpha \; : \; {\pi_\alpha}^{-1}(E) \in \sigma(\mathscr{F}_1) \right\}.$$

Clearly, the set is a σ- algebra containing $\mathscr{E}$ and hence contains $\sigma(\mathscr{E})$. $\square$

The next theorem enables us to use product sets if we have a countable collection. See Appendix for properties of coordinate maps.

Proposition 4.1.5. *If* A *is a countable collection of index sets, then* $\bigotimes_{\alpha \in A} \mathscr{M}_\alpha$ *is the sigma algebra generated by.*

$$\mathscr{F}_2 = \left\{ \prod_{\alpha \in A} E_\alpha \; : \; E_\alpha \in \mathscr{E}_\alpha \right\},$$

where $\mathscr{M}_\alpha = \sigma(\mathscr{E}_\alpha)$*.*

Proof. First we will show that $\bigotimes_{\alpha \in A} \mathscr{M}_\alpha$ is generated by,

$$\mathscr{F} = \left\{ \prod_{\alpha \in A} E_\alpha \; : \; E_\alpha \in \mathscr{M}_\alpha \right\},$$

and then use 4.1.4 to complete the proof. Note that for any $x \in \mathscr{E}$, there is an $E_\alpha \in \mathscr{M}_\alpha$ such that $x = {\pi_\alpha}^{-1}(E_\alpha)$. Now,

$$ {\pi_\alpha}^{-1}(E_\alpha) = \prod_{\beta \in A} Y_\beta, $$

where $Y_\beta = X_\beta$ whenever $\beta \neq \alpha$, and $Y_\beta = E_\alpha$ whenever $\beta = \alpha$. Thus $x \in \mathscr{F}$ i.e, $\mathscr{E} \subset \mathscr{F}$ and hence $\sigma(\mathscr{E}) \subset \sigma(\mathscr{F})$. To show the other inclusion, let $x \in \mathscr{F}$, then $x = \prod\limits_{\alpha \in A} E_\alpha$, where $E_\alpha \in \mathscr{M}_\alpha$. Now,

$$ \prod_{\alpha \in A} E_\alpha = \bigcap_{\alpha \in A} {\pi_\alpha}^{-1}(E_\alpha) \in \sigma(\mathscr{E}). $$

Thus, $\sigma(\mathscr{F}) \subset \sigma(\mathscr{E})$. Now using 4.1.4, we get the result since $\sigma(\mathscr{F}_1) = \sigma(\mathscr{E}) = \sigma(\mathscr{F})$. □

Theorem 4.1.2 (Sigma Algebra in product metric space)**.** Let $X_1, \ldots, X_n$ be metric spaces and let $X = \prod\limits_{j=1}^{n} X_j$,equipped with the product metric. Then $\bigotimes_{j=1}^{n} \mathfrak{B}_{X_j} \subset \mathfrak{B}_X$. If X_j's are **separable**, then $\bigotimes_{j=1}^{n} \mathfrak{B}_{X_j} = \mathfrak{B}_X$.

Proof. □

Corollary 4.1.2.1. $\mathfrak{B}_{\mathbb{R}^n} = \bigotimes_{j=1}^{n} \mathfrak{B}_{\mathbb{R}}$.

The following propositon is useful in considering the measurability of vector valued functions.

Theorem 4.1.3. Let $(X, \mathscr{M})$ and $(Y_\alpha, \mathscr{N}_\alpha)$ $(\alpha \in A)$ be measurable spaces and let $Y = \prod\limits_{\alpha \in A} Y_\alpha$. Define $\mathscr{N}$ to be the product sigma algebra in Y and π_α to be the α^{th} co-ordinate map. Then,

$$ f \;: X \to Y \quad \text{is}\, (\mathscr{M}, \mathscr{N}) - \text{measurable} \quad \text{iff} $$
$$ f_\alpha = \pi_\alpha \circ f \quad \text{is} \quad (\mathscr{M}, \mathscr{N}_\alpha) - \text{measurable for all } \alpha. $$

Proof. By definition, each π_α is $(\mathscr{N}, \mathscr{N}_\alpha)$ – measurable. If f is $(\mathscr{M}, \mathscr{N})$ – measurable then by 4.1.2, f_α is $(\mathscr{M}, \mathscr{N}_\alpha)$ – measurable.

To show that f is measurable given that each f_α is measurable we need to show that for each $E \in \mathscr{E}$, $f^{-1}(E) \in \mathscr{M}$ where $\mathscr{E}$ is the generating set $\mathscr{N}$. Let E be an arbitrary element of $\mathscr{E}$. Then $E = {\pi_\alpha}^{-1}(E_\alpha)$ for some $\alpha \in A$ and $E_\alpha \in \mathscr{N}_\alpha$. Thus,

$$ f^{-1}\left({\pi_\alpha}^{-1}(E_\alpha)\right) = (\pi_\alpha \circ f)^{-1}(E_\alpha) = {f_\alpha}^{-1}(E_\alpha) \in \mathscr{M} $$

since f_α is $(\mathscr{M}, \mathscr{N}_\alpha)$ – measurable. □

Corollary 4.1.3.1. *Let $(X, \mathscr{M})$ be a measurable space. A function $f : X \to \mathbb{R}^m$ is $\mathscr{M}$-measurable iff each co-ordinate function $f_i : X \to \mathbb{R}$, for $1 \le i \le m$, is $\mathscr{M}$-measurable.*

Proof. The function $f : X \to \mathbb{R}^m$ is $\mathscr{M}$-measurable whenever it is $(\mathscr{M}, \mathfrak{B}_{\mathbb{R}^m})-$ measurable. By 4.1.2.1, $\mathfrak{B}_{\mathbb{R}^m} = \bigotimes_{i=1}^m \mathfrak{B}_{\mathbb{R}}$, hence by 4.1.3, f is $\mathscr{M}$-measurable iff $f_i : X \to \mathbb{R}$, for $1 \le i \le m$, is $\mathscr{M}$-measurable. □

Corollary 4.1.3.2. *Let $(X, \mathscr{M})$ be a measurable space. A function $f : X \to \mathbb{C}$ is $\mathscr{M}$-measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathscr{M}$-measurable.*

Proof. By 4.1.3 and the fact that $\mathfrak{B}_{\mathbb{C}} = \mathfrak{B}_{\mathbb{R}} \bigotimes \mathfrak{B}_{\mathbb{R}}$. □

Intuition suggests measurability should not be affected by behavior on sets of measure 0. This is indeed true for complete measures.

Proposition 4.1.6. *Let $(X, \mathscr{M})$, $(Y, \mathscr{N})$ be measure spaces. Let $(X, \mathscr{M})$ be a complete measure space. If f is $(\mathscr{M}, \mathscr{N})$ − measurable and $f = g$ a.e., then g is also $(\mathscr{M}, \mathscr{N})$ − measurable.*

Proof. Let $A = \{x \in X \ : \ f(x) \neq g(x)\}$. By our assumption, measure of A is 0. Fix any $E \subset \mathscr{N}$ and consider the set $g^{-1}(E) = \{x \in X \ : \ g(x) \in E\}$. Then we can write this set a union of disjoint sets as follows,

$$\{x \ : \ g(x) \in E\} = \{x \ : \ g(x) \in E\} \cap A \dot{\bigcup} \{x \ : \ g(x) \in E\} \cap A^{\mathsf{c}}$$
$$\{x \ : \ g(x) \in E\} \cap A \dot{\bigcup} \{x \ : \ f(x) \in E\},$$

where the second set after the equality is due to the fact that $f = g$ in $X - A$. The first set after the equality is a subset of A which has measure 0, since $\mathscr{M}$ is complete and thus is in $\mathscr{M}$. The second set is in $\mathscr{M}$ by our assumption. Thus, $g^{-1}(E)$ is the union of two measurable sets in $\mathscr{M}$ and hence, g is $(\mathscr{M}, \mathscr{N})$ − measurable. □

Before we observe some important properties of measurable functions, let us define the characteristic function of a set. Such a function is the building block for approximating any measurable functions.

Definition 4.1.6 (Characteristic function)**.** *Let $(X, \mathscr{M})$ be a measure space and let $E \subset X$ be any set. The characteristic function $\boldsymbol{\chi}_E$ is given by,*

$$\boldsymbol{\chi}_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \in E^{\mathsf{c}} \end{cases} \tag{4.1}$$

*The characteristic function of a set E is also called the **indicator** function of the set E and is denoted by $\boldsymbol{I}_E$ or $\mathbf{1}_E$.*

It is easy to see (4.1.1 (3)) that $\boldsymbol{\chi}_E$ is measurable if and only if $E \in \mathscr{M}$.

Remark 4.1.4. *Note that if A, B are disjoint sets then $\boldsymbol{\chi}_{A\cup B} = \boldsymbol{\chi}_A + \boldsymbol{\chi}_B$. This means that $\boldsymbol{\chi}_A + \boldsymbol{\chi}_{A^c} = 1$. For any A, B we can observe that $\boldsymbol{\chi}_A \boldsymbol{\chi}_B = \boldsymbol{\chi}_{A\cap B}$. These observations are very useful.*

4.2. Properties of measurable functions

Let $(X, \mathscr{M})$ be a measure space.

Proposition 4.2.1. *If $f : X \to \mathbb{C}$ is $\mathscr{M}$-measurable and $\Phi : X \to C$ is continuous then,*

> ***Property 1*** *$\Phi \circ f$ is $\mathscr{M}$-measurable.*

Proposition 4.2.2. *If $f, g : X \to \mathbb{C}$ are $\mathscr{M}$-measurable then,*

> ***Property 2*** *$f + g$ and fg are $\mathscr{M}$-measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$ by $F(x) = (f(x), g(x))$. Then, since $\mathfrak{B}_{\mathbb{C}\times\mathbb{C}} = \mathfrak{B}_{\mathbb{C}} \bigotimes \mathfrak{B}_{\mathbb{C}}$, F is measurable by 4.1.3. Now, consider the following functions,

- $\phi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ given by, $\phi(z, w) = z + w$,
- $\psi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ given by, $\psi(z, w) = zw$.

Both ϕ, ψ are $(\mathfrak{B}_{\mathbb{C}\times\mathbb{C}}, \mathfrak{B}_{\mathbb{C}})-$measurable, since they are continuous (see 4.1.1.1). Hence, $f + g = \phi \circ F$ and $fg = \psi \circ F$ are $\mathscr{M}$- measurable. $\square$

By taking $g = c$ a constant we see that cf is also $\mathscr{M}$- measurable. If g is function that is not 0, then by using $1/g$ above we see f/g is also $\mathscr{M}$- measurable. Hence, we preserve measurability by doing arithmetic operations.

Proposition 4.2.3. *If (f_j) is a sequence of $\overline{\mathbb{R}}$ valued $\mathscr{M}$- measurable functions then the functions,*

> ***Property 3*** *$\sup\limits_j f_j, \inf\limits_j f_j, \limsup f_j, \liminf f_j$ are $\mathscr{M}$- measurable. If $f = \lim\limits_{n\to\infty} f_j$ exists for all $x \in X$, then f is $\mathscr{M}$- measurable.*

Proof. Let $g_1 = \sup\limits_j f_j$. For each j, $f_j(x) \le g_1(x)$ for all $x \in X$. Let $g > g_1$. Hence, $\{f_j \ge a\} \subset \{g \ge a\}$. Thus for each j,

$$\{f_j \ge a\} \subset \bigcup_{j=1}^{\infty} \{f_j \ge a\} \subset \{g \ge a\}.$$

Hence,

$$\{g_1 \ge a\} = \bigcup_{j=1}^{\infty} \{f_j \ge a\}.$$

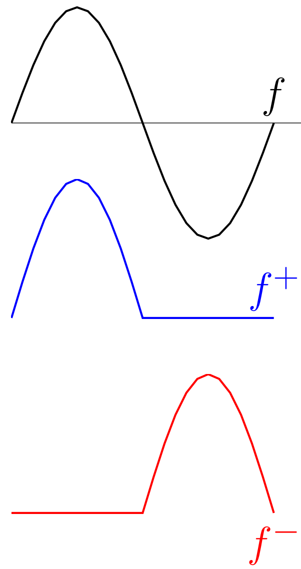

Figure 4.3. Decomposition of a real valued function. Here f^+ is drawn in blue, while f^- is drawn in red.

Thus g_1 is $\mathscr{M}$- measurable since it is a countable union of $\mathscr{M}$- measurable sets. Let $g_2 = \inf\limits_j f_j$. A similar argument yields,

$$\{g_2 \le a\} = \bigcup_{j=1}^{\infty} \{f_j \le a\}.$$

Now, $\liminf f_j = \sup\limits_k \inf \{f_j \ : \ j \ge k\}$ and $\limsup f_j = \inf\limits_k \sup \{f_j \ : \ j \ge k\}$ are $\mathscr{M}$- measurable since $\inf, \sup$ are $\mathscr{M}$- measurable. Finally, if $f = \lim\limits_{j\to\infty} f_j$ exists for all x then $f = \limsup f_j$ and hence f is $\mathscr{M}$- measurable. $\square$

Definition 4.2.1 (Decomposition of real valued functions)**.** *Let $f \ : X \to \overline{R}$. Consider the sets,*

$$A^+ = \{f > 0\} \quad \textit{and} \quad A^- = \{f < 0\}.$$

Then f can be ***decomposed*** *into is positive and negative parts,*

$$f = f^+ - f^-,$$

where $f^+ = f\boldsymbol{\chi}_{A^+} = \max(f, 0)$ and $f^- = -f\boldsymbol{\chi}_{A^-} = \max(-f, 0)$.

Note that each f^+ and f^- is non-negative.

Proposition 4.2.4. *Let $f \ : X \to \overline{R}$.*

Property 4 *f is $\mathscr{M}$- measurable iff f^+, f^- are $\mathscr{M}$- measurable.*

Proof. If f is $\mathscr{M}$- measurable then A^+, A^- are measurable (sets in $\mathscr{M}$) and hence f^+, f^- are $\mathscr{M}$- measurable. If f^+, f^- are $\mathscr{M}$- measurable then f is the sum of two $\mathscr{M}$- measurable functions and so is $\mathscr{M}$- measurable. $\square$

Note that $|f| = f^+ + f^-$. Hence, if f is $\mathscr{M}$- measurable then $|f|$ is also $\mathscr{M}$- measurable.

Proposition 4.2.5. *Let (f_j) be a sequence of $\overline{\mathbb{R}}$ valued $\mathscr{M}$- measurable functions. Let g_1 be a $\mathscr{M}$- measurable function. Let $\mathscr{M}$ be a complete measure space.*

> ***Property 5*** *If $g_1 = g_2$ a.e., then g_2 is $\mathscr{M}$- measurable.*
>
> ***Property 6*** *If $\lim\limits_{j\to\infty} f_j = f$ a.e., then f is $\mathscr{M}$- measurable.*

Proof. By 4.1.6, with $f = g_1$ and $g = g_2$ we get the result.

Let $g = \limsup f_j$. Then by 4.2.3, g is $\mathscr{M}$- measurable. Also $g = f$ a.e., hence using the result from above f is $\mathscr{M}$- measurable. $\square$

4.3. Approximation by simple functions

In this section, we will see that any $\mathscr{M}$- measurable function $f : X \to \mathbb{C}$ can be approximated by simpler functions named appropriately as **simple function**.

Definition 4.3.1. *A simple function $s : X \to \mathbb{C}$ is a finite sum,*

$$s = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where each E_k is a measurable set in $\mathscr{M}$ of finite measure and $a_k \in \mathbb{C}$ are constants. We require that E_k's be disjoint and $X = \dot{\bigcup}_{k=1}^{N} E_k$.

Since the characteristics functions are measurable, a simple function is measurable by 4.2.2. Note that simple functions do not have unique representation as seen in the following example.

Example 4.3.1. *Let $X = [0,3]$ and define*

$$s(x) = \begin{cases} -1, & 0 \le x \le 1 \\ 2, & 1 < x \le 2 \\ 0, & 2 < x \le 3 \end{cases}$$

Let $E_1 = [0,1], E_2 = (1,2], E_3 = (2,3]$. Then,

$$s = -1\chi_{E_1} + 2\chi_{E_2} + 0\chi_{E_3}.$$

However, if we define $D_1 = [0, 1/2), D_2 = [1/2, 1], D_3 = (1, 2], D_4 = (2, 3]$, *then*

$$s = -1\chi_{D_1} - 1\chi_{D_2} + 2\chi_{D_3} + 0\chi_{D_4}.$$

A key result in measure theory is that any function can be approximated by a sequence of monotonic increasing simple functions. We give an alternate characterization for simple functions,

Proposition 4.3.1. *Let* $(X, \mathscr{M})$ *be a measurable space. A function* $s : X \to \mathbb{C}$ *is a simple function if and only if* s *is measurable and the range of* s *is a finite set of points in* $\mathbb{C}$.

Proof. If s is measurable function, then by definition its range is finite. For the other implication, assume range of $s = \{z_i : 1 \leq i \leq n\}$. Let $E_i = s^{-1}(\{z_i\})$. For any $x \in X$, if $x \in E_i$, then $s(x) = z_i$. Thus,

$$s = \sum_{i=1}^{n} z_i \chi_{E_i},$$

is a simple function. Note that $\bigcup_{i=1}^{n} E_i = X$ and E_i's are pairwise disjoint. □

In the proof above we used the **standard** representation of s. It exhibits s as a linear combination, with distinct coefficient, of characteristic functions of disjoint measurable sets whose union is X. A key observation is that the standard representation of a simple function is unique. The proposition that follows shows that the space of simple functions defined on a measure space is a Vector space.

Proposition 4.3.2. *Let* $(X, \mathscr{M})$ *be a measure space.*

> ***Property 1*** *If* s *is a simple function defined on* X *then* cs *is also a simple function.*
>
> ***Property 2*** *If* s_1, s_2 *are simple functions defined on* X *then* $s_1 + s_2$ *is also a simple function.*

Proof. We will show (2). Let $s_1 = \sum_{k=1}^{N} a_k \chi_{E_k}$ and $s_2 = \sum_{j=1}^{M} b_j \chi_{F_j}$ be two simple functions in standard representation. First assume $M = 1$. Note that E_k's and F_1 all partition X and hence $F_1 = X$. Let $G_i = E_i \cap F_1$ for all $1 \leq i \leq N$. Then, $\dot{\bigcup}_{i=1}^{N} G_i = X$ and

$$s_1 + s_2 = \sum_{l=1}^{N} c_l \chi_{G_l},$$

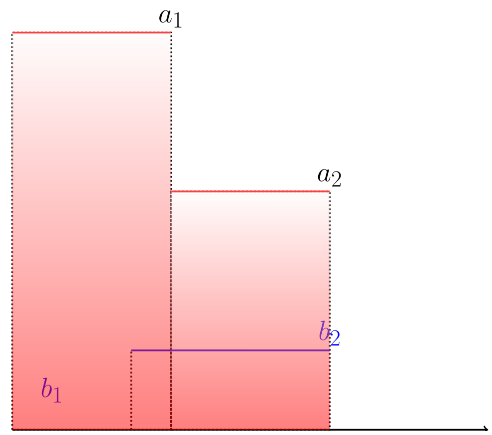

Figure 4.4. Addition of two simple functions.

where $c_l = a_l + b_1$ for $1 \leq l \leq N$. This idea can be extended to all $M \geq 1$ by the using the collection,

$$\{E_k \cap F_j \ : \ 1 \leq k \leq N, 1 \leq j \leq M\},$$

and using the sum $a_k + b_j$ in the corresponding set. See 4.4. Thus we see that,

$$s_1 + s_2 = \sum_{i,j=1} (a_k + b_j) \boldsymbol{\chi}_{E_k \cap F_j}.$$

□

We now come to the most important result in this chapter.

Theorem 4.3.1 (Approximation by simple functions)**.** Let $(X, \mathscr{M})$ be a measure space.

(1) If $f : X \to [0, \infty]$ is measurable, there is a sequence (s_n) of simple functions such that,

$$0 \leq s_1 \leq s_2 \leq s_3 \cdots \leq f \quad \text{and} \quad s_n \to f,$$

pointwise. The convergence is uniform on any set on which f is bounded.

(2) If $f : X \to \mathbb{C}$ is measurable, there is a sequence (s_n) of simple functions such that,

$$0 \leq |s_1| \leq |s_2| \leq |s_3| \cdots \leq |f| \quad \text{and} \quad s_n \to f,$$

pointwise. The convergence is uniform on any set on which f is bounded.

Proof. We prove in order. See 4.5.

(1) The key idea will be to *partition the range* of f. For each n, we partition the co-domain of f by two sets, $[0,n]$ and $(n,\infty]$. We partition $[0,n]$ further and then treat the other set as a *remainder*. In order to get an approximation from below, we need the partition of $[0,n]$ to increase with n which can be done by decreasing the partition length. Thus for each n, we fix a uniform partition length of $\Delta_n = \frac{1}{2^n}$, hence as $n \to \infty$, $\Delta_n \to 0$. Let us index the partition as follows,

$$I_k^n = [k\Delta_n, (k+1)\Delta_n) \quad , 0 \le k < \frac{n}{\Delta_n}.$$

Thus,

$$\begin{aligned} f^{-1}([0,\infty]) &= f^{-1}([0,n]) \bigcup f^{-1}((n,\infty]) \\ &= \bigcup_{k=0}^{\frac{n}{\Delta_n}-1} \left(f^{-1}(I_k^n)\right) \bigcup f^{-1}((n,\infty]) . \end{aligned}$$

Let us set,

$$\begin{aligned} E_k^n &= f^{-1}(I_k^n), \quad 0 \le k < \frac{n}{\Delta_n} = n2^n \\ F_n &= f^{-1}((n,\infty]) \end{aligned}$$

We define the simple function s_n as,

$$s_n = \sum_{k=0}^{n2^n-1} k2^{-n} \boldsymbol{\chi}_{E_k^n} + n\boldsymbol{\chi}_{F_n}.$$

Each I_k^n splits into 2 intervals at the level $n+1$ and hence when $x \in E_k^n$, $s_n(x) \le s_{n+1}(x)$. When $x \in (n, n+1]$, then $s_n(x) = n$ whereas, $s_{n+1}(x) \ge n$. When $x \in (n+1,\infty]$, then $s_n(x) = n$ while $s_{n+1}(x) = n+1$ and thus $s_n \le s_{n+1}$ for all x,n. Also, whenever $f(x) \le N$, then $f(x) \in I_k^n$ for some $k \in 1,\ldots,n2^n$. Thus $f(x) - s_n(x) \le \frac{k+1}{2^n} - \frac{k}{2^n} = \frac{1}{2^n}$, which goes to 0 as $n \to \infty$.

If f is bounded in some set, then there is a N such that $f(x) < N$ for all x in the set. But, for any x, $f(x) - s_N(x) < \frac{1}{2^N}$ and taking N larger and using the fact that $s_n \ge s_N$ for $n \ge N$ we can make $f(x) - s_n(x) < \epsilon$ for any ϵ. Thus, convergence is uniform.

(2) We can decompose $\operatorname{Re} f$, $\operatorname{Im} f$ as f^+, f^- and use the preceeding theorem.

$\square$

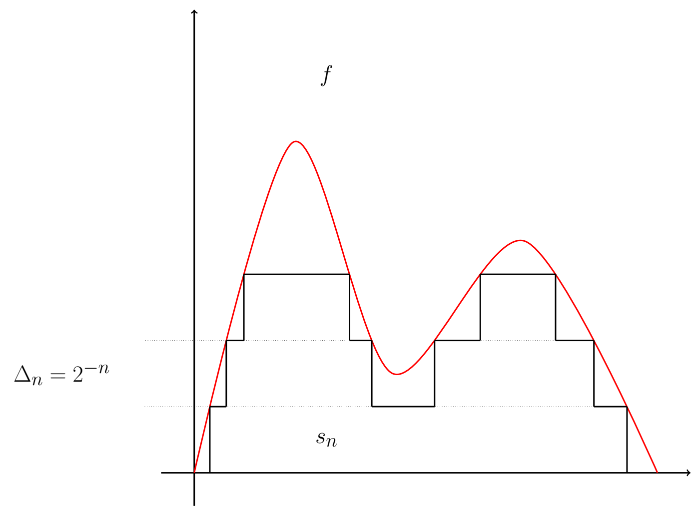

Figure 4.5. Approximation by simple functions

Chapter 5

Integration theory

5.1. Class of Riemann integrable functions

5.2. Abstract integration in measure space

We have seen that Riemann integration doesn't behave nicely when interchanging limit order. To remedy this, we will define the (Lebesgue) integral in a general measure space. In order to achieve this nice property we will expand the functions that can be integrated. As we saw in the previous section, the basic idea of Riemann Integration of a bounded function is to partition the domain. In this section we will partition the range of the function. The theory will proceed in 3 stages:

(1) Define integration for simple functions.

(2) Define integration for non-negative measurable functions.

(3) Define integration for general measurable functions.

In each stage we will show that our integration satisfies the 3 most important property of an integral:

Property 1 Linearity

Property 2 Additivity

Property 3 Monotonicity

We let $(X, \mathscr{M}, \mu)$ be a measure space.

Definition 5.2.1. *If s is a simple function $s = \sum_{k=1}^{N} a_k \boldsymbol{\chi}_{E_k}$, we define the ($\mu$-) integeral of s by,*

$$\int_X s d\mu = \sum_{k=1}^{N} a_k \mu\,(E_k). \tag{5.1}$$

For any $A \in \mathscr{M}$, integral over A is given by,

$$\int_A s d\mu := \int_X s \boldsymbol{\chi}_A d\mu = \sum_{k=1}^{N} a_k \mu\,(E_k \cap A) \tag{5.2}$$

We have seen that a simple functions don't have unique representation and thus we need to check that the above definition is well defined.

Proposition 5.2.1. *Integration of simple function as described by 5.1 is well defined.*

Proof. Suppose $s = \sum_{k=1}^{N} a_k \mu\,(E_k)$ is a simple function where a_k's are not distinct. Note that E_k's are disjoint and their union is X. Let $i = 1$ and j_i be the smallest index of the set $J_1 = \{k \,:\, a_k = a_1\}$ and let $E_{j_1} = \bigcup_{k \in J_1} E_k$ and $a_{j_1} = a_1$. If a_2 is in J_1 we leave it alone and proceed until we find an $i \leq N$ such that $a_i \notin J_1$. Thus, the sets E_{j_i} are disjoint and their union is X. Let N_j be the number of such j_i. Then, $s = \sum_{j_i = j_1}^{N_j} a_{j_i} \boldsymbol{\chi}_{E_{j_i}}$ and hence,

$$\int_X s d\mu = \sum_{j_i = j_1}^{N_j} a_{j_i} \mu\,(E_{j_i}) = \sum_{k=1}^{N} a_k \mu\,(E_k),$$

where the last equality is due to the additivity property of measure. $\square$

Thus we don't have to worry about the representation of the simple function.

Proposition 5.2.2 (Properites of integral of simple functions)**.** *Let $A \in \mathscr{M}$. The integral of simple functions defined in 5.1 satisfies the following properties:*

Property 1 *(Linearity) If s_1, s_2 are simple functions and $a, b \in \mathbb{R}$ then*

$$\int_A (as_1 + bs_2) d\mu = a \int_A s_1 d\mu + b \int_A s_2 d\mu.$$

Property 2 *(Additivity) Let s_1 be a simple function. If A, B are disjoint subsets of $\mathscr{M}$ then,*

$$\int_{A\cup B} s_1 d\mu = \int_A s_1 d\mu + \int_B s_1 d\mu.$$

Property 3 *(Monotonicity) Let s_1, s_2 be two simple functions such that $s_1 \leq s_2$. Then,*

$$\int_A s_1 d\mu \leq \int_A s_2 d\mu.$$

Property 4 *(Triangle Inequality) Let s_1 be a simple function, then,*

$$\left|\int_A s_1 d\mu\right| \leq \int_A |s_1| d\mu.$$

Proof. We prove in order.

- Assume $a, b = 1$. Let $s_1 = \sum_{i=1}^{N_1} a_i \chi_{E_i}$ and $s_2 = \sum_{j=1}^{N_2} b_j \chi_{F_j}$ We know that by 4.3.2,

  $$s_1 + s_2 = \sum_{i,j=1} (a_i + b_j)\chi_{E_i \cap F_j}.$$

  Hence,

  $$\begin{aligned}
  \int_A (s_1 + s_2) d\mu &= \sum_{i,j=1} (a_i + b_j)\mu\left(E_i \cap F_j \cap A\right) \\
  &= \sum_{i,j=1} a_i \mu\left(E_i \cap F_j \cap A\right) + \sum_{i,j=1} b_j \mu\left(E_i \cap F_j \cap A\right) \\
  &= \sum_i a_i \sum_j \mu(E_i \cap F_j \cap A) + \sum_j b_j \sum_i \mu(E_i \cap F_j \cap A) \\
  &= \sum_i a_i \mu\left(E_i \cap A\right) + \sum_j b_j \mu\left(F_j \cap A\right) \\
  &= \int_A s_1 d\mu + \int_A s_2 d\mu.
  \end{aligned}$$

  The equality in the 4^{th} line is due to the fact that E_i's and F_j's *partition* X and hence $\{E_i \cap A\}$ and $\{F_j \cap A\}$ are partitions of A, so we use the additivity of measure. It is easy to observe for any a, b different than 1.

- Noting that measure is additive for disjoint sets we can observe,

$$\begin{aligned}\int\limits_{A\cup B} s_1 d\mu &= \sum_{i=1}^{N} a_i\mu\left(E_i \cap (A\cup B)\right)\\ &= \sum_{i=1}^{N} a_i\mu\left((E_i\cap A)\cup(E_i\cup B)\right)\\ &= \sum_{i=1}^{N} \left(a_i\mu\left(E_i\cap A\right)\right)+\left(a_i\mu\left(E_i\cup B\right)\right)\\ &= \sum_{i=1}^{N} \left(a_i\mu\left(E_i\cap A\right)\right)+\sum_{i=1}^{N}\left(a_i\mu\left(E_i\cup B\right)\right)\\ &= \int\limits_A s_1 d\mu + \int\limits_B s_1 d\mu\end{aligned}$$

- If $s \geq 0$ is a simple function, then all its coefficient are greater than equal to 0 and so $\int\limits_X s d\mu \geq 0$. Thus if $s = s_2 - s_2$, then $s \geq 0$ and we get the result.
- Let $s_1 = \sum\limits_{i=1}^{N_1} a_i \boldsymbol{\chi}_{E_i}$. We get the result by observing that,

$$\left|\sum_{i=1}^{N} a_i\mu\left(E_i\right)\right| \leq \sum_{i=1}^{N} \left|a_i\right|\mu\left(E_i\right).$$

□

Monotonicity shows that the Integral is a **positive operator** on the space of simple functions i.e. it maps positive simple functions to non-negative real numbers.

Remark 5.2.1. *Note that if s_1, s_2 are simple functions such that $s_1 = s_2$ a.e. then $\int\limits_X s_1 d\mu = \int\limits_X s_2 d\mu$ since the difference is 0 except on a set of measure zero, the integral of the difference is 0 and hence we can use linearity to reach to the conclusion.*

If A is any set of measure 0, then $\int\limits_A s d\mu$ must be 0 provided we have a complete measure space. To see this, note that $\int\limits_A s d\mu = \sum\limits_{k=1}^{N} a_k\mu\left(E_k \cap A\right)$. Since A is a measure 0, and $E_k \cap A \subset A$, we can see that if we have a complete measure space $\mu\left(E_k\cap A\right) = 0$ and hence $\int\limits_A s d\mu = 0$. Since any measure space can be completed we are not too concerned about completeness.

We now embark on stage 2; defining integral of non-negative measurable functions.

Definition 5.2.2. *Let $f : X \to [0, \infty]$ be a non-negative measurable function. Then, we define,*

$$\int_X f d\mu = \sup \left\{ \int_X s d\mu \ : \ 0 \leq s \leq f, \ s \text{ is simple} \right\}. \tag{5.3}$$

Note that the integral can be infinite. We say that f is (μ-) integrable if $\int_X f d\mu < \infty$.

Note that because of 4.3.1 (1), the above definition makes sense. We can always approximate any measurable function by a sequence of simple functions from below and hence f will be given by the limit of such a sequence (which will be the sup). For any $A \subset \mathscr{M}$, we define $\int_A f d\mu := \int_X f \boldsymbol{\chi}_A d\mu$.

Proposition 5.2.3 (Properties of the integral of non-negative measurable function)**.** *Let $A \in \mathscr{M}$. Let f, g be a non-negative extended real valued function defined on X such that its integral is given by 5.3. Then, it satifies the following properties:*

> ***Property 1*** *(Linearity) For any $a, b \in \mathbb{R}$,*
>
> $$\int_A (af + bg) d\mu = a \int_A f d\mu + b \int_A g d\mu.$$
>
> ***Property 2*** *(Additivity) If A, B are disjoint subsets of $\mathscr{M}$ then,*
>
> $$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$
>
> ***Property 3*** *(Monotonicity) Let $f \leq g$. Then,*
>
> $$\int_A f d\mu \leq \int_A g d\mu.$$

At this point only monotonicity is immediately observable from the definition.

Proof. (Monotonicity) First let us consider the case when $A = X$. If $f = g$ the result is obvious. Assume $f < g$. Approximate f, g from below by sequence of non-negative simple functions. Then there will be a simple function s such that $f < s \leq g$. Thus the collection of simple functions approximating g contains the collection of simple functions approximating f. Hence taking the supremum of the integrals we get,

$$\int_X f d\mu \leq \int_X g d\mu.$$

For any $A \in \mathscr{M}$, $f \leq g$ implies $f\boldsymbol{\chi}_A \leq g\boldsymbol{\chi}_A$ and hence we get the result by using the result on $A = X$ above. $\square$

Remark 5.2.2. *Linearity and additivity are not so obvious. We will need some additional results to be able to prove it. One such result is very useful and is called the Monotone Convergence theorem (MCT). Note that one direction of linearity is obvious. If we have simple functions s_f, s_g approximating f, g, then the sequence of $s_f + s_g$ will approximate $f + g$. Since f, g are non-negative, the collection of $s_f + s_g$ will contain s_f and s_g and so by monotonicity $\int_X (f+g)d\mu \geq \int_X f d\mu + \int_X g d\mu$. The other inequality isn't apparent. Note that for any $a \in \mathbb{R}$, $\int_X a f d\mu$ is equal to $a \int_X f d\mu$. This is obvious from the definition.*

Theorem 5.2.1 (Monotone convergence theorem (MCT))**.** If (f_n) is a sequence of monotone increasing non-negative extended real valued measurable functions,

$$0 \leq f_1 \leq f_2 \ldots f_n \leq f_{n+1} \leq \ldots f,$$

and

$$\lim_{n\to\infty} f_n = f,$$

then,

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu.$$

Proof. Let $I_n = \int_X f_n d\mu$ and $I = \int_X f d\mu$. We will show that,

$$\lim_{n\to\infty} I_n \leq I,$$

and

$$\lim_{n\to\infty} I_n \geq I.$$

By monotonicity, for any n we have,

$$I_n \leq I_{n+1} \leq I.$$

Thus the sequence (I_n) is monotone increasing and is bounded by I. Hence, $\lim_{n\to\infty} I_n$ exists and is equal to $\sup \{I_n \ : \ n \in \mathbb{Z}^+\}$ which is less than or equal to I. Thus,

$$\lim_{n\to\infty} I_n \leq I.$$

To prove the other inequality let s be a simple function such that $0 \leq s \leq f$. This is possible due to 4.3.1. Consider the set,

$$A_n = \{x \in X \ : \ f_n(x) \geq ts(x)\},$$

where $t \in (0,1)$ is a real number. Note that (A_n) is an *increasing* sequence of sets that increase to X i.e.

$$A_n \nearrow X.$$

To see why this is true, all we need to check that if $x \in X$ is an arbitrary element of X, then it must be in atleast one of the sets A_n. If $f(x) = 0$ then each $f_n = 0$ and so x is in all A_n. If $f(x) > 0$, then there must be n far out such that $x \in A_n$. This is because

$$0 < ts(x) < s(x) \leq f(x),$$

hence there must be a f_n such that

$$0 < ts(x) < f_n(x) \leq f(x),$$

because $f_n \to f$ for any $x \in X$. This is the reason for taking a $t \in (0,1)$. Now because $f \geq f\boldsymbol{\chi}_{A_n}$, by monotonicity,

$$\int_X f_n d\mu \geq \int_{A_n} f_n d\mu \geq \int_{A_n} tsd\mu = t\int_{A_n} sd\mu.$$

Let us write s in standard representation,

$$s = \sum_{k=1}^{N} a_k \boldsymbol{\chi}_{E_k},$$

for $E_k \in \mathscr{M}$. Then,

$$\begin{aligned}\lim_{n\to\infty} \int_{A_n} sd\mu &= \lim_{n\to\infty} \sum_{k=1}^{N} a_k \mu\left(E_k \cap A_n\right) \\ &= \sum_{k=1}^{N} a_k \lim_{n\to\infty} \mu\left(E_k \cap A_n\right) \\ &= \sum_{k=1}^{N} a_k \mu\left(E_k \cap X\right) \\ &= \int_X sd\mu\end{aligned}$$

We used the fact that $\mu(A_n) \to \mu(X)$ because $A_n \nearrow X$. Hence,

$$\lim_{n\to\infty} \int_X f_n d\mu \geq t\int_X sd\mu.$$

Since $t \in (0,1)$ was arbitrary we must have,

$$\lim_{n\to\infty} \int_X f_n d\mu \geq \int_X s d\mu.$$

This means that $\lim\limits_{n\to\infty} I_n$ is an upper bound for the set,

$$\left\{ \int_X s d\mu \ : \ 0 \leq s \leq f, \ s \text{ is simple} \right\},$$

and hence is greater than equal to the supremum of the set which is by definition equal to $\int\limits_X f d\mu = I$. Hence,

$$\lim_{n\to\infty} I_n \geq I.$$

$\square$

The monotone convergence theorem (MCT) is a very useful theorem. Immediately we can see that in the case where we have a monotone increasing non-negative real valued function, we can *bring* the limit **inside** the integral. We will explore some important convergence theorem later. Now, we will use MCT to show linearity.

Proof. (Linearity). Let (s_n), (t_n) be non-negative simple functions such that $s_n \nearrow f$ and $t_n \nearrow g$. Hence $as_n + bt_n \nearrow af + bg$. Thus,

$$\begin{aligned}
\int_X (af+bg) d\mu &= \lim_{n\to\infty} \int_X (as_n + bt_n) d\mu \\
&= \lim_{n\to\infty} \left(\int_X as_n d\mu + \int_X bt_n d\mu \right) \\
&= \lim_{n\to\infty} \int_X as_n d\mu + \lim_{n\to\infty} \int_X bt_n d\mu \\
&= \int_X af d\mu + \int_X bg d\mu,
\end{aligned}$$

where the first and last equality is from 5.2.1 (MCT), and the second inequality is due to linearity of the integral w.r.t simple functions. If A is any set in $\mathscr{M}$, then $s_n \boldsymbol{\chi}_A \nearrow f\boldsymbol{\chi}_A$ and $t_n\boldsymbol{\chi}_A \nearrow g\boldsymbol{\chi}_A$. Hence $(as_n + bt_n)\boldsymbol{\chi}_A \nearrow (af+bg)\boldsymbol{\chi}_A$. Thus we can use the above result. $\square$

Although additivity can now follow from what we have proved, we will prove additivity by showing that a non-negative measurable function induces a measure. First, a useful lemma

Lemma 5.2.1. *Assume* $f_1, f_2, \ldots$ *are non-negative measurable functions defined on* X*. Then,*

$$\int_X \sum_{k=1}^{\infty} f_k d\mu = \sum_{k=1}^{\infty} \int_X f_k d\mu.$$

Proof. Let $F_n = \sum_{k=1}^{n} f_k$ and let $F = \sum_{k=1}^{\infty} f_k$. Then, since f_k's are non-negative we have

$$0 \leq F_1 \leq F_2 \leq \ldots F.$$

Each F_n is measurable and $F_n \nearrow F$ and hence F is measurable. Thus we can use MCT to see that

$$\begin{aligned}\int_X F d\mu &= \lim_{n\to\infty} \int_X F_n d\mu \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \int_X f_k d\mu \\ &= \sum_{k=1}^{\infty} \int_X f_k d\mu,\end{aligned}$$

where the second equality is due to linearity of the integral. □

Lemma 5.2.2. *If* f *is a non-negative extended real valued measurable function and* $E \in \mathscr{M}$ *be a set such that* $\mu(E) = 0$*, then*

$$\int_E f d\mu = 0.$$

Proof. Note that $\int_E f d\mu = \int_X f \boldsymbol{\chi}_E d\mu$. If (s_n) is a sequence of non-negative simple functions such that $s_n \nearrow f$, then $s_n \boldsymbol{\chi}_E \nearrow f \boldsymbol{\chi}_E$. But we have seen that (see 5.2.1) $\int_X s_n \boldsymbol{\chi}_E d\mu = \int_E s_n d\mu = 0$ and hence we get the result. □

Next we come to an important statement.

Theorem 5.2.2. Let f be a non-negative extended real valued function. Then,

$$\mu_f(E) := \int_E f d\mu,$$

for each $E \in \mathscr{M}$ defines a measure on $(X, \mathscr{M})$.

Proof. From the Lemma above it is evident that $\mu_f(\emptyset) = 0$. Now let $E = \bigcup\limits_{k=1}^{\infty} E_k$ where the E_k's are pairwise disjoint sets in $\mathscr{M}$. Note that $\boldsymbol{\chi}_E = \sum\limits_{k=1}^{\infty} \boldsymbol{\chi}_{E_k}$ since the E_k's are disjoint. Then,

$$\begin{aligned}
\mu_f(E) &= \int\limits_E f d\mu \\
&= \int\limits_X f \boldsymbol{\chi}_E d\mu \\
&= \int\limits_X \sum_{k=1}^{\infty} f \boldsymbol{\chi}_{E_k} d\mu \\
&= \sum_{k=1}^{\infty} \int\limits_X f \boldsymbol{\chi}_{E_k} d\mu \\
&= \sum_{k=1}^{\infty} \int\limits_{E_k} f d\mu \\
&= \sum_{k=1}^{\infty} \mu_f(E_k),
\end{aligned}$$

where the 4^{th} inequality is from the Lemma following the monotone convergence theorem. Hence μ_f is a measure on $(X, \mathscr{M})$. □

Corollary 5.2.2.1 (Additivity)**.** *Additivity of the integral.*

Proof. (Additivity) If A, B are disjoint sets in $\mathscr{M}$, then $\mu_f(A \cup B) = \mu_f(A) + \mu_f(B)$ i.e. $\int\limits_{A \cup B} f d\mu = \int\limits_A f d\mu + \int\limits_B f d\mu$. □

Corollary 5.2.2.2. *If $A \subset B$, then $\int\limits_A f d\mu \leq \int\limits_B f d\mu$.*

Proof. This follows from the monotononcity of μ_f that $\mu_f(A) \leq \mu_f(B)$. □

Now we embark on the final stage (3) of our construction. If f is an extended real valued measurable function, we have seen that we can decompose f in two non-negative parts,

$$f = f^+ - f^-.$$

We have already seen how to define the integral for f^+, f^-. However, our definition allows these integrals to be infinite and hence we must be careful about things like $\infty - \infty$. Note that $|f| = f^+ + f^-$.

Definition 5.2.3. *If $f\ : X \to [-\infty, \infty]$ is a measurable function, then*

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu, \tag{5.4}$$

provided atleast one of the integral on the right hand side is not infinite. We say that f is integrable if both $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ are finite. Note that, this is equivalent to saying $\int_X |f| d\mu < \infty$.

If $A \in \mathscr{M}$ and $\int_X f d\mu$ is defined then,

$$\int_A f d\mu := \int_X f \boldsymbol{\chi}_A d\mu.$$

Proposition 5.2.4 (Properties of the integral of non-negative measurable function)**.** *Let $A \in \mathscr{M}$. Let f, g be extended real valued functions defined on X such that their integral is given by 5.4. Then, it satifies the following properties:*

Property 1 *(Linearity) For any $a, b \in \mathbb{R}$,*

$$\int_A (af + bg) d\mu = a \int_A f d\mu + b \int_A g d\mu.$$

Property 2 *(Additivity) If A, B are disjoint sets in $\mathscr{M}$ then,*

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

Property 3 *(Monotonicity) Let $f \leq g$. Then,*

$$\int_A f d\mu \leq \int_A g d\mu.$$

Property 4 *(Triangle Inequality)*

$$\left| \int_A f d\mu \right| \leq \int_A |f| d\mu$$

Proof. We prove in order.

- We prove for $A = X$. The result for any $A \in \mathscr{M}$ can then be proved by using $f\chi_A$ and $g\chi_A$ instead of f, g. We will prove the results for integrable function i.e. we assume the integrals of both f^+, f^- and g^+, g^- are finite. The general proof can then be completed by taking all the different cases.

  We will show that $\int\limits_X af d\mu = a \int\limits_X f d\mu$ and $\int\limits_X (f+g) d\mu = \int\limits_X f d\mu + \int\limits_X g d\mu$.

  If $a > 0$, then $(af)^+ = af^+$ and $(af)^- = af^-$. $\int\limits_X af d\mu$ is $\int\limits_X (af)^+ d\mu - \int\limits_X (af)^- d\mu$ and thus,

$$\int\limits_X af d\mu = a \int\limits_X f^+ d\mu - a \int\limits_X f^- d\mu = a \int\limits_X f d\mu.$$

  If $a < 0$ then $(af)^+ = -af^-$ and $(af)^- = -af^+$ and the result follows analogously.

  Let $h = f + g$. Then $h^+ - h^- = f^+ - f^- + g^+ - g^-$. Hence,

$$h^+ + f^- + g^- = h^- + f^+ + g^+.$$

  Thus, using the linearity of non-negative functions we get,

$$\int\limits_X h^+ d\mu + \int\limits_X f^- d\mu + \int\limits_X g^- d\mu = \int\limits_X h^- d\mu + \int\limits_X f^+ d\mu + \int\limits_X g^+ d\mu.$$

  Since we have assumed each of them are finite, we get

$$\int\limits_X h^+ d\mu - \int\limits_X h^- d\mu = \int\limits_X f^+ d\mu - \int\limits_X f^- d\mu + \int\limits_X g^+ d\mu - \int\limits_X g^- d\mu.$$

  Hence we get the result.
- For any set $E \in \mathscr{M}$, $(f\chi_E)^+ = f^+\chi_E$ and $(f\chi_E)^- = f^-\chi_E$. Thus using $E = A \cup B$,

$$\int\limits_{A\cup B} f d\mu = \int\limits_X f\chi_{A\cup B} d\mu = \int\limits_X f^+\chi_{A\cup B} d\mu - \int\limits_X f^-\chi_{A\cup B} d\mu.$$

  Now from additvity of non-negative functions,

$$\int\limits_X f^+\chi_{A\cup B} d\mu = \int\limits_X f^+\chi_A d\mu + \int\limits_X f^+\chi_B d\mu,$$

  and

$$\int\limits_X f^-\chi_{A\cup B} d\mu = \int\limits_X f^-\chi_A d\mu + \int\limits_X f^-\chi_B d\mu.$$

Hence, we get the result.

- Assume f is integrable. Since $f \leq g$, $0 \leq g - f$ and thus using linearity of integrals of non-negative functions we get the result. To complete the proof we just need to check all the cases on the decompositons of f, g.
- This follows from the fact that $-|f| \leq f \leq |f|$ and monotonicity above.

$\square$

An integral of a measurable function can tell us something about the function. More precisely if an integral is finite it must be the case that the function is finite. As with any case in measure theory, we should make a deduction about properties a.e. rather than pointwise.

Proposition 5.2.5. *If $f : X \to [-\infty, \infty]$ is an integrable function then f is finite a.e. on X.*

Proof. Let $E = \{x \in X : |f(x)| = \infty\}$. Take any $t > 0$ and consider the function $t\boldsymbol{\chi}_E$. By definition $|f| > t\boldsymbol{\chi}_E$ and hence from monotonicity $\int\limits_X |f| d\mu \geq \int\limits_X t\boldsymbol{\chi}_E d\mu = t\mu(E)$. If $\mu(E)$ is finite, this would mean the $\int\limits_X |f| d\mu > t$ for any t which would contradict the hypothesis that f is integrable. Hence $\mu(E) = 0$ which means that both f^+, f^- are finite a.e. on X and hence f is finite a.e. on X $\square$

Proposition 5.2.6. *Suppose that $f : X \to [-\infty, \infty]$ is a measurable function. Then,*

$$\int\limits_X |f| d\mu = 0 \iff f = 0 \quad a.e.$$

Proof. Note that if $f = 0$ a.e. then $|f| = 0$ a.e. and vice versa. Hence, we can just replace f by $|f|$ in the Proposition above without changing anything. This just amounts to assuming that f is positive.

If f is a simple function $f = \sum\limits_{k=1}^{N} a_k \boldsymbol{\chi}_{E_k}$ such that $f = 0$ a.e. then it means that either all its coefficient a_k's are zero or the measure of its component sets E_k's is zero and hence $\int\limits_X f d\mu = 0$. If f is not a simple function, then for any simple function $0 \leq s \leq f$ implies $s = 0$ a.e. and hence $\int\limits_X f d\mu = 0$ by definition of the integral of an non-negative measurable function.

Now let $\int\limits_X |f| d\mu = 0$. Let E be the set where f is not equal to 0. Since f is assumed positive, this means that E is the set $\{f > 0\}$. Let us define the following set,

$$E_n = \left\{ x \in X \ : \ f(x) \geq \frac{1}{n} \right\}.$$

Fix any $\epsilon > 0$. There is an $n \in \mathbb{Z}^+$ such that $1/n < \epsilon$ and hence if $f(x) > \epsilon$ then $f(x) > 1/n$. This means that,

$$E \subset \bigcup_{n=1}^{\infty} E_n,$$

Note that $0 \leq \frac{1}{n}\boldsymbol{\chi}_{E_n} \leq f$ for all n by construction and hence,

$$\frac{1}{n}\mu\left(E_n\right) = \int\limits_X \frac{1}{n}\boldsymbol{\chi}_{E_n} d\mu \leq \int\limits_X f d\mu = 0.$$

Thus,

$$\mu\left(E\right) \leq \sum_{n=1}^{\infty} \mu\left(E_n\right) \leq 0.$$

Thus E has 0 measure. □

Intuitively, integrable functions should in some sense vanish at infinity since their integrals are finite. We make precise this notion in the following theorem.

Theorem 5.2.3 (Absolute continuity)**.** Suppose f is integrable extended real valued function on X. Then for any $\epsilon > 0$

(1) There is a set of finite measure $E \in \mathscr{M}$ such that $\int\limits_{E^c} |f| d\mu < \epsilon$.

(2) There is a $\delta > 0$ such that for any $E \in \mathscr{M}$,

$$\int\limits_E |f| d\mu < \epsilon,$$

whenever $\mu\left(E\right) < \delta$.

Proof. We prove in order. By replacing f with $|f|$, we don't change anything. Hence we can assume f to be non-negative.

(1) Let $E_n = \{x \in X \ : \ 0 \leq f(x) \leq n\}$. Then $E_1 \subset E_2 \subset \ldots$. Let $f_n = f\boldsymbol{\chi}_{E_n}$. Then $f_n \nearrow f$. Hence we can use MCT to state that

$$\lim_{n\to\infty} \int\limits_X f_n d\mu = \int\limits_X f d\mu.$$

Thus there is an N such that,

$$\left|\int_X f_n d\mu - \int_X f d\mu\right| < \epsilon,$$

whenever $n \geq N$. Since $f \geq f_n$ for all n, we have

$$\int_X f d\mu - \int_X f_N d\mu < \epsilon,$$

which from linearity can be written as,

$$\int_X (f - f\boldsymbol{\chi}_{E_N}) d\mu < \epsilon.$$

Since $1 - \boldsymbol{\chi}_{E_N} = \boldsymbol{\chi}_{E_N^{\mathsf{c}}}$ we get,

$$\epsilon > \int_X f\boldsymbol{\chi}_{E_N^{\mathsf{c}}} d\mu = \int_{E_N^{\mathsf{c}}} f d\mu.$$

E_N is of finite measure by 5.2.5. Hence, we get the result.

(2) With the same definition for f_n as above we can observe that,

$$\begin{aligned}\int_E f d\mu &= \int_E (f - f_n) d\mu + \int_E f_n d\mu \\ &\leq \int_X (f - f_n) d\mu + n\mu(E),\end{aligned}$$

for any $n \in \mathbb{Z}^+$ and $E \in \mathscr{M}$. The first term in the second line can be made smaller than $\frac{\epsilon}{2}$ by MCT for some $N \in \mathbb{Z}^+$. Pick $\delta = \frac{\epsilon}{2N}$. Hence when $\mu(E) < \delta$,

$$\int_E f d\mu \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus, we get the result.

□

5.3. Limit theorems

One of the basic questions in analysis is when can we interchange limit operations? Note that integration is a limit operation (the supremum of simple functions is the limit of the simple functions that approximate a given non-negative measurable function). If $f_n \to f$ pointwise, when can one say that

$$\int_X f_n d\mu \to \int_X f d\mu?$$

The monotone convergence theorem says that the above is true if f_n increases to f for non-negative measurable functions. In a general case we may not have monontone sequences. The simplest case in which we can guarantee the convergence of the integral is when $f_n \to f$ **uniformly**. To see this, note that

$$\int_X f_n d\mu - \int_X f d\mu \leq \int_X |f_n - f| d\mu \leq \mu(X)\,\epsilon,$$

for n greater than some N. An important caveat is that here we have adopted the convenction that

$$\infty \cdot 0 = 0 \cdot \infty = 0.$$

This certainly works for Riemann integral, however uniform convergence is too strong a demand. One of the advantages of the construction of lebesgue integral is that we can (almost) acheive the convergence of integrals when we have pointwise convergence of functions. A few cautionary examples follow,

Example 5.3.1. *Let $(\mathbb{R}, \mathscr{L}, \mu)$ be the space of lebesgue measure. Let $f_n : \mathbb{R} \to \mathbb{R}$ be given by,*

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ pointwise but $\int_{\mathbb{R}} f_n d\mu = 1$. (Note that f_n is a simple function and hence we were able to calculate it easily.)

Example 5.3.2. *Let $(\mathbb{R}, \mathscr{L}, \mu)$ be the space of lebesgue measure. Let $f_n : \mathbb{R} \to \mathbb{R}$ be given by,*

$$f_n(x) = \begin{cases} n^2 & \text{if } 0 < x < \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f_n \to 0$ pointwise but $\int_{\mathbb{R}} f_n d\mu = n$ and hence $\int_{\mathbb{R}} f_n d\mu \to \infty$ (Note that f_n is a simple function and hence we were able to calculate it easily.)

Thus we need some extra assumptions to deduce convergence of integrals. The first important theorem generalizes the monotonce convergence theorem for non-negative functions.

Theorem 5.3.1 (Fatou's Lemma)**.** Suppose that (f_n) is a sequence of positive measurable functions $f_n \ : X \to [0, \infty]$. Then

(5.5)
$$\int_X \liminf_n f_n d\mu \leq \liminf_n \int_X f_n d\mu.$$

Proof. Let $g_n = \inf_{k \geq n} f_k$. Then $g_n \nearrow \liminf_n f_n$. Hence we can use MCT to deduce that,

$$\lim_{n\to\infty} \int_X g_n d\mu = \int_X \liminf_n f_n d\mu.$$

Now each $g_n \leq f_k$ for $k \geq n$ and hence,

$$\int_X g_n d\mu \leq \int_X f_k d\mu,$$

for $k \geq n$. Hence,

$$\int_X g_n d\mu \leq \inf_{k\geq n} \int_X f_k d\mu,$$

and taking $n \to \infty$ we see that,

$$\lim_{n\to\infty} \int_X g_n d\mu \leq \liminf_n \int_X f_n d\mu.$$

Hence, we get the result. For any set $E \in \mathscr{M}$, multiplying f_n, f by $\boldsymbol{\chi}_E$ we can get the result for the integral over E. $\square$

Now we come to the most important result in this chapter which gives us the conditions when to expect the convergence of the integral of the functions given pointwise convergence of the functions.

Theorem 5.3.2 (Dominated Convergence theorem)**.** Let $E \in \mathscr{M}$ be any measurable set and let (f_n) be a sequence of measurable functions such that for each n, $|f_n| \leq g$ a.e. on E where g is an integrable function over E. If $f_n \to f$ a.e. on E then f is integrable over E and,

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu.$$

Proof. Since $-g < f_n < g$ we have $g + f_n > g - g = 0$ and hence $g + f_n$ is a non-negative function such that $g + f_n \to f$. Note that $\liminf_n f_n = f$.

Using Fatou's lemma on $g + f_n$ we get

$$\begin{aligned}\int_E g d\mu + \int_E f d\mu &= \int_E (g+f) d\mu \\ &= \int_E \liminf_n (g+f_n) d\mu \\ &\leq \liminf_n \int_E (g+f_n) d\mu, \\ &= \int_E g d\mu + \liminf_n \int_E f_n d\mu.\end{aligned}$$

Subtracting $\int_E g d\mu$ since g is integrable we get,

$$\int_E f d\mu \leq \liminf_n \int_E f_n d\mu.$$

If we show

$$\int_E f d\mu \geq \limsup_n \int_E f_n d\mu,$$

we are done. Note that $(g - f_n) > 0$ and thus using the same argument we get,

$$\int_E g d\mu - \int_E f d\mu \leq \int_E g d\mu + \liminf_n (-\int_E f_n d\mu).$$

Since $\liminf_n (-a_n) = -\limsup_n a_n$ we get

$$\int_E f d\mu \geq \limsup_n \int_E f_n d\mu.$$

Hence, we get the result. $\square$

5.4. Relation with Riemann Integral

5.5. Product spaces and Fubini's theorem

Let $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ be two measure spaces. We have already seen that the product sigma algebra is defined by,

$$\mathscr{M} \otimes \mathscr{N} = \sigma\left(\{A \times B \ : \ A \in \mathscr{M}, B \in \mathscr{N}\}\right).$$

Definition 5.5.1. *Let $\mathscr{E} = \{A \times B \ : \ A \in \mathscr{M}, B \in \mathscr{N}\}$. Then $\mathscr{E}$ is called the collection of measurable rectangles. By definition, $\mathscr{M} \otimes \mathscr{N} = \sigma(\mathscr{E})$.*

Our goal in this section would be introduce a measure Π on the sigma algebra $\mathscr{M} \otimes \mathscr{N}$, such that,

$$\Pi(A \times B) = \mu(A)\nu(B).$$

Note that the sigma algebra contains sets E which cannot be written as a product of sets $A \times B$ for some $A \in \mathscr{M}$ and $B \in \mathscr{N}$. Let us introduce some terminology and notation.

Definition 5.5.2 (sections)**.** *Suppose that X, Y are sets and that $E \subset X \times Y$. Then for each $x \in X$ and each $y \in Y$ the sections E_x and E^y are the subsets defined as,*

$$E_x = \{y \in Y \ : \ (x, y) \in E\},$$
$$E^y = \{x \in X \ : \ (x, y) \in E\}.$$

If f is a function defined on $X \times Y$, the sections f_x, f^y are the functions on Y and X respectively given by, $f_x(y) = f(x, y)$ and $f^y(x) = f(x, y)$.

Lemma 5.5.1. *Let $(X, \mathscr{M})$ and $(Y, \mathscr{N})$ be measure spaces.*

(1) If $E \subset X \times Y$ belongs to the product sigma algebra $\mathscr{M} \otimes \mathscr{N}$, then, E_x belongs to $\mathscr{N}$ and E^y belongs to $\mathscr{M}$.

(2) If f is an extended real valued $\mathscr{M} \otimes \mathscr{N}$-measurable function defined on $X \times Y$, then each section f_x is $\mathscr{N}$-measurable and each section f^y is $\mathscr{N}$-measurable.

Proof. We prove in order.

(1) Fix an $x \in X$. Let

$$\mathscr{F} = \{E \in \mathscr{M} \otimes \mathscr{N} \ : \ E_x \in \mathscr{N}\}.$$

Note that $\mathscr{F} \subset \mathscr{M} \otimes \mathscr{N}$. If we show that $\mathscr{F} \supset \mathscr{E}$ and $\mathscr{F}$ is a sigma algebra, we are done. If $E \in \mathscr{E}$, then $E = A \times B$ for some A, B in $\mathscr{M}$ and $\mathscr{N}$ respectively. It is easy to see that E_x is either B or $\emptyset$; and in both cases $E_x \in \mathscr{N}$. Thus, $\mathscr{E} \subset \mathscr{F}$. To show that $\mathscr{F}$ is a sigma algebra, we need to show that it is closed under complements and countable unions. Note that, $(E_x)^{\mathsf{c}} = (E^{\mathsf{c}})_x$. Thus, if $E \in \mathscr{F}$, then since $E \in \mathscr{M} \otimes \mathscr{N}$ which is a sigma-algebra, $E^{\mathsf{c}} \in \mathscr{M} \otimes \mathscr{N}$. Moreover, since $E_x \in \mathscr{N}$ by our hypothesis, $(E_x)^{\mathsf{c}}$ is in $\mathscr{N}$ since $\mathscr{N}$ is a sigma-algebra. Similarly for countable unions. Since this is true for any $x \in X$, we get the result. The same arguments can be used to show that E^y belongs to $\mathscr{M}$.

(2) Note that,

$$(f^{-1}(D)) = \{(x, y) \ : \ f(x, y) \in D\}.$$

We show that ${f_x}^{-1}(D) = \left(f^{-1}(D)\right)_x$ i.e. fix an $x \in X$ and observe that,

$$\begin{aligned}\left(f^{-1}(D)\right)_x &= \{y \;:\; f(x,y) \in D\},\\ &= \{y \;:\; f_x(y) \in D\},\\ &= {f_x}^{-1}(D).\end{aligned}$$

For any $D \in \mathfrak{B}_{\mathbb{R}}$, $f^{-1}(D) \in \mathscr{M} \otimes \mathscr{N}$ and by the previous result, $\left(f^{-1}(D)\right)_x \in \mathscr{N}$. Thus, ${f_x}^{-1}(D) \in \mathscr{N}$ and so is $\mathscr{N}$-measurable.

□

Proposition 5.5.1. *Let $(X, \mathscr{M}, \mu)$, $(X, \mathscr{N}, \nu)$ be σ-finite measure spaces. If E belongs to the product sigma-algebra $\mathscr{M} \otimes \mathscr{N}$, then the function that maps x to $\nu(E_x)$ is $\mathscr{M}$-measurable and the function that maps y to $\mu(E^y)$ is $\mathscr{B}$-measurable.*

Proof. We will prove the Proposition in steps for the case of finite measure.

(1) We show that the family of measurable rectangles $\mathscr{E}$ is a π-class. If $E_1 = A_1 \times B_1$ and $E_2 = A_2 \times B_2$, then

$$E_1 \cap E_2 = (A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2).$$

Thus, $E_1 \cap E_2 \in \mathscr{E}$.

(2) We show that a sub-collection $\mathscr{F}$ of $\mathscr{M} \otimes \mathscr{N}$ satisfying the hypothesis in the Proposition is a λ-class. Let $h \;: X \to \mathbb{R}$ be given by $h(x) = \nu(E_x)$. By the above Lemma, such a function is well-defined. Similarly, let $l \;: Y \to \mathbb{R}$ be given by $l(y) = \mu(E^y)$. Define,

$$\mathscr{F} = \{E \in \mathscr{M} \otimes \mathscr{N} \;:\; x \mapsto \nu(E_x), y \mapsto \mu(E_y) \text{ are measurable}\}.$$

Let $E = X \times Y$. Then, $E_x = Y$ and $E^y = X$. Consider $h(x) = \nu(E_x)$. Then, $h^{-1}((a, \infty])$ is either $\emptyset$ or X; in both the cases $h^{-1}((a, \infty]) \in \mathscr{M}$. Thus, h is $\mathscr{M}$-measurable for $E = X \times Y$ and by similar reasonings l is $\mathscr{N}$-measurable. Let $E \in \mathscr{F}$. Let $h(x) := \nu((E_x)^{\mathsf{c}})$. Let $g(x) = \nu(Y)$ and $h'(x) = \nu(E_x)$. Then, by our hypothesis $h'(x)$ is measurable. For any $x \in X$, $\nu((E_x)^{\mathsf{c}}) = \nu(Y) - \nu(E_x)$ and hence h is a sum of two measurable functions and is thus measurable. Thus, $E^{\mathsf{c}} \in \mathscr{F}$. If (E_n) is a sequence of pairwise disjoint sets in $\mathscr{F}$, then for any x

$$\nu\Big(\bigcup_{n=1}^{\infty} (E_n)_x\Big) = \lim_{n \to \infty} \sum_{i=1}^{n} \nu((E_i)_x).$$

Thus, $\bigcup\limits_{n=1}^{\infty} E_n \in \mathscr{F}$, because the limit of measurable functions is also measurable, i.e. , let $h'_n = \sum\limits_{i=1}^{n} h_i$, where $h_i(x) = \nu((E_i)_x)$. Then, each h_i is measurable by our hypothesis, and so h'_n is measurable for each n and hence h is measurable (being the limit of measurable functions.) Hence, $\mathscr{F}$ is a λ-class.

(3) We appeal to $\pi - \lambda$ Theorem to conclude that $\mathscr{F} = \mathscr{M} \otimes \mathscr{N}$. By our construction, $\mathscr{F} \subset \mathscr{M} \otimes \mathscr{N}$. Also we showed that $\mathscr{F} \supset \mathscr{E}$ and $\mathscr{F}$ is a λ-class. Since, $\mathscr{E}$ is a π-class, by the $\pi - \lambda$ Theorem, $\mathscr{F} \supset \sigma(\mathscr{E}) = \mathscr{M} \otimes \mathscr{N}$. Hence, $\mathscr{F} = \mathscr{M} \otimes \mathscr{N}$.

We can extend this for the case when μ, ν are sigma finite. In other words, we only need to show Step (2) when μ, ν are sigma-finite. But this is easily seen by applying to the case of $\nu(E_x \cap Y_n)$, where $Y_n \nearrow Y$ and $\nu(Y_n) < \infty$ and analogously for μ.

□

Theorem 5.5.1. Let $(X, \mathscr{M}, \mu)$, $(Y, \mathscr{F}, \nu)$ be σ-finite measure spaces. Then, there is a unique measure Π on the sigma-algebra $\mathscr{M} \otimes \mathscr{N}$, such that,

$$\Pi(A \times B) = \mu(A)\nu(B),$$

for every $A \in \mathscr{M}$ and $B \in \mathscr{N}$. Furthermore, for any arbitrary set $E \in \mathscr{M} \otimes \mathscr{N}$,

$$\Pi(E) = \int\limits_X \nu(E_x) d\mu = \int\limits_Y \mu(E^y) d\nu.$$

The measure Π is the called the product measure and is sometimes denoted by $(\mu \times \nu)$.

Proof. For any $E \in \mathscr{M} \otimes \mathscr{N}$, define,

$$\Pi_1(E) := \int\limits_X \nu(E_x) d\mu,$$

$$\Pi_2(E) := \int\limits_Y \mu(E_y) d\nu$$

By the previous Theorem, $x \mapsto \nu(E_x)$ is a $\mathscr{M}$-measurable function and so Π_1 is well-defined. Similarly Π_2 is well-defined. We show that Π_1, Π_2 are measures. Clearly $\Pi_1(\emptyset) = 0$. Let (E_n) be a sequence of measurable sets in $\mathscr{M} \otimes \mathscr{N}$. Then, $(E_x)_n$ is a sequence of measurable sets in $\mathscr{N}$. Let

$E = \bigcup_{n=1}^{\infty} E_n$.

$$\begin{aligned}\Pi_1(E) &= \int_X \nu(\bigcup_{n=1}^{\infty} (E_x)_n) d\mu \\ &= \int_X \sum_{n=1}^{\infty} \nu((E_x)_n) d\mu \\ &= \sum_{n=1}^{\infty} \int_X \nu((E_x)_n) d\mu \\ &= \sum_{n=1}^{\infty} \Pi_1(E_n).\end{aligned}$$

Similarly, Π_2 is measure on $\mathscr{M} \otimes \mathscr{N}$. For any $E \in \mathscr{E}$, where $\mathscr{E}$ is the family of measurable rectangles, we can write $E = A \times B$ for some $A \in \mathscr{M}$ and $B \in \mathscr{N}$. Then,

$$\int_X \nu(B) \boldsymbol{\chi}_A d\mu = \nu(B)\mu(A) = \int_Y \mu(A) \boldsymbol{\chi}_B d\nu.$$

Hence, $\Pi_1 = \Pi_2 = \Pi$ on $\mathscr{E}$ which is a π-class. Thus by 3.6.4, these two measure agree on $\sigma(\mathscr{E}) = \mathscr{M} \otimes \mathscr{N}$. $\square$

WE can extend this to any finite collection of measurable spaces. The following two theorems enable one to evaluate integrals with respect to product measures by evaluating iterated integrals.

Proposition 5.5.2 (Tonelli's Theorem)**.** *Let $(X, \mathscr{M}, \mu)$,$(Y, \mathscr{N}, \nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [0, \infty]$ be $\mathscr{M} \otimes \mathscr{N}$-measurable. Then,*

(1) the functions $x \mapsto \int_Y f_x d\nu$ and $y \mapsto \int_X f^y d\mu$ are $\mathscr{M}, \mathscr{N}$ measurable respectively and

(2) f satisfies

$$\int_{X \times Y} f d\Pi = \int_X \left(\int_Y f_x d\nu \right) d\mu = \int_Y \left(\int_X f^y d\mu \right) d\nu.$$

Proof. For both (1) and (2), we will show the result holds for characteristic functions and simple functions. And then by MCT, the result holds for f. For (1), when f is a characteristic function, we get the result from Proposition 5.5.1. The result then follows for simple functions by linearity and for general f by MCT .

For (2), when f is a characteristic function, we get the result from 5.5.1. The result then follows for simple functions by linearity and for general f by MCT . $\square$

Note that Tonelli's theorem is valid for any extended non-negative function that is $\mathscr{M} \otimes \mathscr{N}$ measurable and not necessarily Π integrable. When f is allowed to be negative the result is no longer valid and we need an extra assumption that f be Π integrable. This is Fubini's theorem.

Theorem 5.5.2 (Fubini's Theorem)**.** Let $(X, \mathscr{M}, \mu)$,$(Y, \mathscr{N}, \nu)$ be σ-finite measure spaces, and let $f : X \times Y \to [-\infty, \infty]$ be $\mathscr{M} \otimes \mathscr{N}$-measurable. Then,

(1) For (μ)-a.e. $x \in X$, the section f_x is ν-integrable and for (ν)-a.e. $y \in Y$, the section f^y is μ-integrable,

(2) the functions I_f, J_f defined by,

$$I_f(x) = \begin{cases} \int\limits_Y f_x d\nu & \text{if } f_x \text{ is } \nu\text{-integrable,} \\ 0 & \text{otherwise,} \end{cases}$$

and,

$$J_f(y) = \begin{cases} \int\limits_X f^y d\mu & \text{if } f^y \text{ is } \mu\text{-integrable,} \\ 0 & \text{otherwise,} \end{cases}$$

are μ-integrable and ν-integrable respectively.

(3) the relation,

$$\int\limits_{X \times Y} f d\Pi = \int\limits_X I_f d\mu = \int\limits_Y J_f d\nu,$$

holds.

Proof. By Tonelli's theorem,

$$\int\limits_{X \times Y} |f| d\Pi = \int\limits_X \left(\int\limits_Y |f_x| d\nu \right) d\mu = \int\limits_Y \left(\int\limits_X |f^y| d\mu \right) d\nu.$$

Since, the LHS is finite by hypothesis, each of the *inner* integral is finite. Again by Tonell's theorem, $x \mapsto \int\limits_\nu f_x^+ dY$ and $x \mapsto \int\limits_\nu f_x^- dY$ are $\mathscr{M}$ measurable and by the first statement above are μ-integrable. This means that they are finite for a.e. $x \in X$. Thus, f_x is ν integrable for a.e. $x \in X$, which means that I_f is μ-integrable. By linearity

$$\begin{aligned}\int\limits_{X\times Y} f d\Pi &= \int\limits_{X\times Y} f^+ d\Pi - \int\limits_{X\times Y} f^- d\Pi \\ &= \int\limits_X \left(\int\limits_Y (f^+)_x d\nu\right) d\mu - \int\limits_X \left(\int\limits_Y (f^-)_x d\nu\right) d\mu \\ &= \int\limits_X I_f d\mu.\end{aligned}$$

The same argument applies to J_f.

$\square$

Chapter 6

Spaces of integrable functions

Part 3

Probability Theory

Chapter 7

Probability space

We specialize the general measure theory to provide a description of a probability space.

Definition 7.0.1 (Probability space)**.** *A probability space is a measure space $(X, \mathscr{M}, \mu)$ with $\mu(X) = 1$. The measure μ is called a probability measure. The element of $\mathscr{M}$ are called events. In classical notation, X is denoted by Ω and is called the sample space, $\mathscr{M}$ is denoted by $\mathscr{F}$ and μ by $\mathbb{P}$. Thus a probability space is denoted by $(\Omega, \mathscr{F}, \mathbb{P})$.*

Note that if $E \in \mathscr{F}$ then $E^{\mathsf{c}} \in \mathscr{F}$. This means that $\mathbb{P}(E^{\mathsf{c}}) = 1 - \mathbb{P}(E)$. This is a very useful result that is used very frequently in probability.

Example 7.0.1. *We consider a few probability spaces.*

(1) Let Ω be countable set, $\Omega = \{\omega_1, \omega_2, \dots\}$. Suppose each point ω_i has probability p_i of occuring so that $\sum_i p_i = 1$. Let $\mathscr{F} = \mathscr{P}(X)$. Define,

$$\mathbb{P}(A) = \sum_{x_i \in A} p_i, \quad A \in \mathscr{F}.$$

Then, $\mathbb{P}$ is a probability measure. $(\Omega, \mathscr{F}, \mathbb{P})$ is called the discrete probability space.

(2) Let $\Omega = [0,1]$, $\mathscr{F}$ be the lebesgue measurable sets in $[0,1]$ and $\mathbb{P}$ be the Lebesgue measure. This is called the uniform distribution.

(3) Let $\Omega = \mathbb{R}$ and set $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$. The function $F(x)$ is continuous, monotone increasing and non-negative with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. The distribution function F defines a unique measure on $\mathfrak{B}_{\mathbb{R}}$. This is called the Cauchy distribution.

Next, we describe a measure theoretic probability model.

Definition 7.0.2 (Measure theoretic probability model)**.** *Let X be the space of a probabilistic process. A measure theoretic model of the process is an identification of X with a set Ω, a σ- algebra $\mathscr{F}$ on Ω and a measure $\mathbb{P}$ on $\mathscr{F}$. A subset $E \subset X$ is called a* plausible *event if $\Omega_E \in \mathscr{F}$, where Ω_E is the set of points in Ω which corresponds to points in X for which E occurs. We denote the probability of E by $P(E)$ and set it to $\mathbb{P}(\Omega_E)$.*

The notation Ω_E here signifies that X could be any probabilistic set up where we want to assign probabilities to subsets E of X. In order to build a measure theoretic framework we identify with X a set which we call the sample space Ω and construct a σ- algebra on Ω. Such an identification and construction is not unique. To assign probabilities to events E in X, we see if the corresponding set is in $\mathscr{F}$. If it is, we assign it a probability; if not, we deem the set E as not plausible within our measure theoretic framework.

Now we will build a measure theoretic probability model to analyze an experiment that involves a sequence of infinite coin tosses. Let us assume we have a fair coin and we toss it infinitely many times. Also it will be safe to assume that any outcome of a toss doesn't depend on what preceeded it and will not affect any future outcomes. Ofcourse, such an experiment is just a thought experiment and cannot be practically carried out. Our sample space will consist of *points* like $(H, T, H, H, H, T, H, T, T, H, H, \ldots,)$. We may have events such as the event E of all those infinite sequences such that the first toss is a H. How do we assign a probability to such an event. First a couple of definitions.

Definition 7.0.3 (Bernoulli trial and sequences.)**.** *Suppose an experiment has two possible outcomes. A finite number of repetitions of the experiment is a called a Bernoulli trial. An infinite sequence of experiments is called a Bernoulli sequence.*

Let $\mathcal{B}$ be the set of all Bernoulli sequences. Thus $\mathcal{B}$ is the space of our experiment of infinite coin tosses. In order to prescribe probability to events in a measure theoretic framework we need to identify $\mathcal{B}$ with the sample space of a probabilistic space. The following proposition says that there is a one to one correspondence between the unit interval and $\mathcal{B}$.

Proposition 7.0.1. *There exists a $1-1$ correspondence from $(0,1]$ into $\mathcal{B}$.*

Proof. Let $\omega \in [0,1]$. Then ω can be written as a binary expansion given by,

$$\omega = \sum_{i=1}^{\infty} \frac{a_i}{2^i}, \quad a_i = 0 \,\mathrm{or}\, 1.$$

Let $f\ : [0,1] \to \mathcal{B}$ be the map given by,

$$f(\omega) = \begin{cases} H & \text{if } a_i = 1, \\ T & \text{if } a_i = 0. \end{cases}$$

The map defined above fails to be a function since there are real numbers that have two different binary expansion. For example $\frac{1}{2} = 0.10\ldots$ or $\frac{1}{2} = 0.0111\ldots$. To avoid this trouble we chose the later expansion. What this means is that the function fails to be onto since we discard binary expansions terminating in an infinite sequences of 0 which corresponds to Bernoulli sequence ending in infinite T's. Thus there is a $1-1$ correspondence from $(0,1]$ into $\mathcal{B}$. $\square$

Remark 7.0.1. *Since $(0,1]$ is uncountable, this implies that $\mathcal{B}$ is uncountable. However, because the map isn't bijective we cannot identify every bernoulli sequence with a real number in the unit interval. However there is only a countable subset of $\mathcal{B}$ that is not mapped by f above and since we know that a measure of a countable set is 0 we can neglect this set to build our measure theoretic framework. To show this, let $\mathcal{B}_{neg}$ be the Bernoulli sequence that are not mapped by f. These are the sequences that end in infnite T's. Let $\mathcal{B}^k_{neg}$ be the bernoulli sequences that end in inifinite T's after the k^{th} toss. Then $\mathcal{B}^k_{neg}$ is countable. But $\mathcal{B}_{neg} = \bigcup_k \mathcal{B}^k_{neg}$ and so the set that is not mapped by f is countable.*

We can make $\mathcal{B}$ into a probability space by using the Lebesgue measure on the unit interval.

Definition 7.0.4 (Borel Principle)**.** *Let $\Omega = (0,1]$. If E is a **plausible** event of Bernoulli sequences, we denote by Ω_E as the subset of real numbers in $(0,1]$ that are Lebesgue measurable and are given by the corresponding binary expansion. We set the probability of E, denoted by $P(E)$, as $\mu(\Omega_E)$.*

Example 7.0.2. *Let E = event where H occurs on the first toss. $E = \{H, X_1, X_2, X_3, \ldots\}$. The corresponding set in Ω is*

$$\Omega_E = \{\omega \in \Omega; x = 0.1d_1d_2d_3\ldots\ :\ d_i = 0\, or 1\}\,.$$

Is E plausibe? That is is Ω_E Lebesgue measurable? The smallest number in Ω_E is $0.100000\ldots$ while the largest number is $0.11111\ldots$. Thus $\Omega_E = (1/2,1]$ which is certainly Lebesgue measurable. Thus $P(E) = \mu((1/2,1]) = 1/2$.

Example 7.0.3. *Let E be the event where the first N tosses are prescribed. For example we could have for $N=3$, HTT as the first 3 tosses. Thus,*

$$\Omega_E = \{\omega \in \Omega; x = 0.a_1a_2\ldots a_Nd_1d_2d_3\ldots\ :\ d_i = 0\, or 1\}\,.$$

Is Ω_E Lebesgue measurable? Again the smallest number is $a_1a_2\dots a_N00000\dots$ while the largest number is $a_1a_2\dots a_N11111$. Let $s = a_1a_2\dots a_N000\dots$. Then $a_1a_2\dots a_N111\dots = s + \sum_{i\geq N+1}^{\infty} 1/2^i = s = 1/2^N$. Thus $\Omega_E = (s, s+\frac{1}{2^N}]$ which is certainly Lebesgue measurable. Thus $P(E) = \mu\left((s, s+\frac{1}{2^N}]\right) = \frac{1}{2^N}$.

Example 7.0.4. *Consider the event E in which H occurs in the N^{th} toss. The corresponding set*

$$\Omega_E = \{\omega \in \Omega; x = 0.d_1d_2\dots 1d_{N+1}d_{N+2}d_{N+3}\dots \ : \ d_i = 0\, or 1\}$$

Here Ω_E can be thought of as a union of disjoint interval. As a concrete case, consider $N = 3$. Then we have the following cases: HTH, HHH, THH, HTH. These correspond to 4 disjoint intervals of length $1/2^3$. Thus $P(E) = 4/8 = 1/2$. In general, $P(E) = 2^{N-1}/2^N = 1/2$.

In all these examples the Events have corresponded to either intervals or union of intervals. However, the power of a measure theoretic framework is not apparent since the above examples can be easily seen even in a discrete setting. To show the power of the Borel principle, we will look at the Weak Law of Large Numbers. See Appendix for a connection between weak law of large numbers and the approximation of continuous functions by polynomials.

We believe that we should be able to detect the probabilities of Heads and Tails when flipping a coin by examining the results of many experiments. For example suppose the probability that Head occurs in a coin toss is p. If we don't know what p is we can carry the experiment a large number of times and record the number of Heads that occured. The fraction of Heads then must give us some indication of p. However, a precise statement of this intuition is difficult to formulate.

Say we are tossing a fair coin whose results are independent of any other coin tosses. If we let H_n to be the number of heads that occur in the first n tosses, we **might** like to show

$$\lim_{n\to\infty} \frac{H_n}{n} = \frac{1}{2}.$$

However this is certainly not true. In other words, we could get a sequence that ends in H infinitely or in T infinitely. Intuitively, such sequences are not likely or typical. What is clear is that we need to create a careful formulation.

Definition 7.0.5 (Number of Heads)**.** *For $\omega \in \Omega$, define*

$$S_N(\omega) = a_1 + \cdots + a_N, \quad \omega = 0.a_1a_2\ldots$$

S_N gives the number of heads in the first N tosses of the Bernoulli sequence corresponding to ω. So $H_N = S_N(\omega)$.

Now, consider the event E that the fraction of heads H tends to 1/2. The corresponding set

$$\Omega_E = \left\{ \omega \in \Omega \; : \; \frac{S_n(\omega)}{n} \to \frac{1}{2} \quad n \to \infty \right\} \tag{7.1}$$

Is Ω_E Lebesgue measurable? This set is highly complicated. What can we say about its elements? If $\omega \in \Omega_E$ then no matter what ϵ we take, there is an integer k such that for all integers n greater than k, the fraction of heads is within an ϵ of $\frac{1}{2}$. That is,

$$\forall\,(\epsilon)\,\exists\,(k)\,\forall\,(n)\left[n \geq k \implies \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| < \epsilon \right].$$

By the archimedes principle, for every $\epsilon > 0$, there is a positive integer r such that $\frac{1}{r} < \epsilon$. Thus the above condition can be written as,

$$\forall\,(r)\,\exists\,(k)\,\forall\,(n)\left[n \geq k \implies \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| < \frac{1}{r} \right].$$

The advantage of such a formulation is that now we have positive integers as quantifiers. Let,

$$A_{n,r} = \left\{ \omega \in \Omega \; : \; \left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| < \frac{1}{r} \right\}.$$

This means that,

$$\frac{n}{2}(\frac{1}{2} - \frac{1}{r}) < a_1 + a_2 + \cdots + a_n < \frac{n}{2}(\frac{1}{2} + \frac{1}{r}).$$

Since a_i are either 0 or 1 we have a finite choice of these. Thus $A_{n,r}$ is a disjoint union of intervals in Ω and is Lebesgue measurable. This means that,

$$\Omega_E = \bigcap_{r=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n \geq k}^{\infty} A_{n,r},$$

is Lebesgue measurable since it is just a countable union and intersection of a Lebesgue measurable set. This is easy to see: if $\omega \in \Omega_E$ then for every r, $\omega \in \bigcup_{k=1}^{\infty} \bigcap_{n \geq k}^{\infty} A_{n,r}$. This means that there is a k such that $\omega \in \bigcap_{n \geq k}^{\infty} A_{n,r}$. Thus for all $n \geq k$, $\omega \in A_{n,r}$, i.e $\left| \frac{S_n(\omega)}{n} - \frac{1}{2} \right| < \frac{1}{r}$. Thus Ω_E is Lebesgue measurable in Ω.

What is $P(E)$ i.e. what is $\mu(\Omega_E)$? Intiuitely, we would guess that $P(E) = 1$. Indeed that is the case and the result is the famous **Strong Law of Large numbers**. To prove this statement we will show $\mu(\Omega_E{}^{\mathsf{c}}) = 0$. There are two ways to show a set has zero Lebesgue measure. Either the set is countable or it can be covered by intervals whose countable sum of measures can be made arbitrarily small. See 2.0.6. We can rule out the first option which seems surprising at first but follows from a simple observation. Consider the map, $\sigma\ : \Omega \to \Omega$ given by,

$$\sigma(\omega) = a_1 11 a_2 11 a_3 11 \cdots, \quad \omega = a_1 a_2 a_3 \cdots.$$

Since σ is injective, $\sigma(\Omega) \subset \Omega$ is uncountable. Fix an $\omega \in \Omega$. Let $N = 3n$ for any positive integer n. Let $y \in \Omega$ be give by $\sigma(\omega)$. Then $S_N(y)$ is the number of heads in the first N tosses which is always greater than $2n$, thus

$$\frac{S_N(y)}{N} \geq \frac{2}{3}.$$

This means that $\sigma(\Omega) \subset \mathscr{F}_E^{\mathsf{c}}$ and hence $\mathscr{F}_E^{\mathsf{c}}$ is uncountable.

But first we will show a *weaker* statement. The notions weaker and stronger will be explained later when we describe measurable functions. Given an $\epsilon > 0$ let

$$\Omega_{E_N} = \left\{ \omega \in \Omega \ : \ \left| \frac{S_n(\omega)}{N} - \frac{1}{2} \right| \leq \epsilon \right\}. \tag{7.2}$$

Note that, Ω_{E_N} is Lebesgue measurable because we can re-formulate the definition by using $\frac{1}{r}$, where r is a positive integer such that $\frac{1}{r} < \epsilon$.

We gather these two notions in the following theorem.

Theorem 7.0.1 (Law of large numbers for Bernoulli sequences)**.** Let $\Omega = (0, 1]$ be the sample space identified with the Bernoulli sequences. Let $\mathscr{F}$ be the Lebesgue measurable sets in Ω and μ be the corresponding Lebesgue measure. Thus, $(\Omega, \mathscr{F}, \mu)$ is a probability space. Let Ω_E be the set in $\mathscr{F}$ as in 7.1 and Ω_{E_N} be the set in $\mathscr{F}$ as in 7.2. Then,

(1) For every $\epsilon > 0$,

$$\lim_{N\to\infty} \mu\left(\left\{ \omega \in \Omega \ : \ \left| \frac{S_n(\omega)}{N} - \frac{1}{2} \right| \leq \epsilon \right\} \right) = 1$$

i.e.,

$$P(E_N) \to 1 \,\text{as}\, N \to \infty$$

(2)

$$\mu\left(\left\{ \omega \in \Omega \ : \ \lim_{N\to\infty} \frac{S_N(\omega)}{N} = \frac{1}{2} \right\} \right) = 1$$

i.e.,

$$P(E) = 1.$$

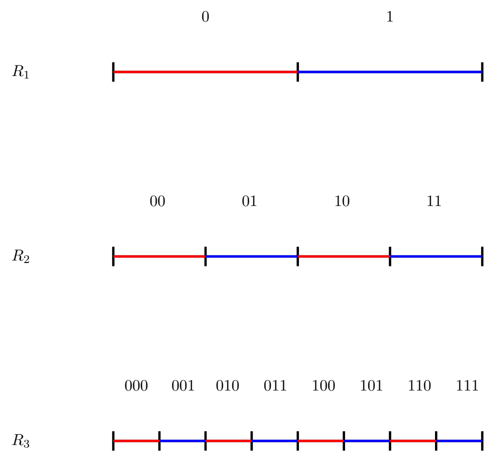

Figure 7.1. We show the first three levels of dyadic partition with the corresponding Rademacher functions R_1, R_2 and R_3. Red shows that Rademacher functions takes a value -1, while blue shows that the function takes a value 1.

We will need a few more definitions and results from Riemann integration to prove this. The standard approach will be to prove results about the complement of the set.

Each digit of the binary expansion of a real number in Ω (where we choose a unique binary expansion) can be thought as a selection of a dyadic partition of Ω as in 2.0.4. At level 1 we bisect the interval $(0, 1]$. The binary digit when in left is 0 and 1 when in right. At level 2, we bisect each of these to get four intervals indexed as $00, 01, 10, 12$. Continuing, we achieve a dyadic partition of $(0, 1]$ at each level which corresponds to a digit in the binary expansion. Since each digit corresponds to a Bernoulli trial, we get a correspondence of the partition with the Bernoulli sequence.

Each Level k partitions Ω into 2^k intervals of $(\frac{l}{2^k}, \frac{l+1}{2^k}]$ for $0 \leq l < 2^k$. At each level of the partition we define the Rademacher functions which take the value of 1 when the binary digit is 1 and -1 when the binary digit is 0. Thus R_1 has two values, R_2 has 4 values, R_n has 2^n values. This is made precise in the definition below. See 7.1.

Definition 7.0.6 (Rademacher function)**.** *For $\omega \in \Omega$, we define the k^{th} Rademacher function by,*

$$R_k(\omega) = 2a_k - 1, \quad \omega = 0.a_1a_2\ldots$$

This means that,

$$R_k(\omega) = \begin{cases} 1, & a_k = 1, \\ -1, & a_k = 0 \end{cases}$$

We can interpret R_k like this. Suppose we bet on a sequence of coin tosses such that at each toss, we win \$1 if it is heads and lose \$1 if it is tails. Then $R_k(\omega)$ is the amount won or lost at the k^{th} toss in the sequence of tosses represented by ω.

Proposition 7.0.2. *Let R_k denote the k^{th} Rademacher function.*

(1) $\int\limits_0^1 R_{k_1}(x)R_{k_2}(x)\dots R_{k_n}(x) = 0\, or\, 1$ *for any sequence* $k_1 \leq k_2 \leq \dots \leq k_n$.

(2) For $x \in (0,1]$,

$$\sum_{i=1}^{\infty} R_i(x)2^{-i} = 2x-1.$$

Proof. We prove in order.

(1) Each R_{k_i} is defined on partitions of $(0,1]$ into 2^{k_i} intervals of $(\frac{l}{2^{k_i}}, \frac{l+1}{2^{k_i}}]$ for $0 \leq l < 2^{k_i}$. Let the partition at level k_n be given by

$$\left\{ \Delta_l^{k_n} = (\frac{l}{2^{k_n}}, \frac{l+1}{2^{k_n}}] \ : \ 0 \leq l < 2^{k_n} \right\}.$$

If $k_1 < k_2 < \dots < k_n$, R_{k_i} for $k_i < k_n$ will be constant on each of these $\Delta_l^{k_n}$. Thus,

$$\begin{aligned}\int\limits_0^1 R_{k_1}(x)\dots R_{k_n}(x) &= \sum_{l=0}^{2^{k_n}-1} \int\limits_{\Delta_l^{k_n}} R_{k_1}(x)\dots R_{k_n}(x) \\ &= (R_{k_1}(x)\dots R_{k_{n-1}}(x)) \int\limits_{\Delta_0^{k_n}+\Delta_1^{k_n}} R_{k_n}(x) \\ &+ \dots + (R_{k_1}(x)\dots R_{k_{n-1}}(x)) \int\limits_{\Delta_{2^{k_n}-1}^{k_n}+\Delta_{2^{k_n}}^{k_n}} R_{k_n}(x)\end{aligned}$$

Each of the integral is 0 and hence the sum is 0. If $k_1 = k_2 = k_3 \dots = k_n$ and n is even then total integral is one otherwise integral is 0.

(2) Let us write $R_i(x) = 2 * a_i - 1$ where $x = 0.a_1a_2\cdots$ is the binary representation of x. Note that $x = \sum_{i=1}^{\infty} a_i/2^i$.

$$\begin{aligned}\sum_{i=1}^{\infty} R_i(x)2^{-i} &= \lim_{n\to\infty}\sum_{i=1}^{n} R_i(x)2^{-i}\\ &= \lim_{n\to\infty}(2*(a_1/2 + a_2/4 + \cdots + a_n/2^n))\\ &\quad - \lim_{n\to\infty}(1/2 + 1/4 + \cdots + 1/2^n)\\ &= 2*x-1\end{aligned}$$

We get the second term by using the definition of R_i, i.e $R_i(x) = 2 * a_i - 1$. We can split the limit because each converges. The first limit is just the binomial representation of x and the second is a geometric series.

□

Definition 7.0.7. *We define* $W_N(\omega) = \sum_{k=1}^{N} R_k(\omega)$. *Then* W *gives the total amount won or lost after the* N^{th} *toss. Using the definition of* R_k*, we get*

$$\begin{aligned}W_N(\omega) &= 2(a_1 + a_2 + \cdots + a_N) - N\\ &= 2S_N(\omega) - N.\end{aligned}$$

With the above definition, our condition in the set given by 7.2 becomes,

$$|W_N(\omega)| \leq 2\epsilon N.$$

We will need one more result before proving 7.0.1. The special case of this result will be proved later.

Proposition 7.0.3. *Let* f *be a non-negative, piecewise constant function on* Ω *and* $\alpha > 0 \in \mathbb{R}$ *be a positive real number. Then,*

$$\mu\left(\{\omega \in \Omega \ : \ f(\omega) > \alpha\}\right) < \frac{1}{\alpha}\int_0^1 f(\omega),$$

where the integeral is the standard Riemann integral in $\mathbb{R}$.

Proof. Suppose f is defined on the mesh $0 = \omega_1 < \omega_2 < \cdots < \omega_k = 1$ such that $f(\omega) = c_i$ whenever $\omega \in (\omega_i, \omega_{i+1})$ for all $1 \le i \le k-1$. Then,

$$\begin{aligned}
\int_0^1 f(\omega) &= \sum_{i=1}^{k} c_i(\omega_{i+1} - \omega_i) \\
&\ge \sum_{\substack{i=1 \\ c_i > \alpha}}^{k} c_i(\omega_{i+1} - \omega_i) \\
&> \alpha \sum_{\substack{i=1 \\ c_i > \alpha}}^{k} (\omega_{i+1} - \omega_i) \\
&= \alpha\mu\left(\{\omega \in \Omega \;:\; f(\omega) > \alpha\}\right).
\end{aligned}$$

□

Now we are ready to prove 7.0.1 (1) which is the **WLLN** for Bernoulli sequences.

Proof. The complement of the set Ω_{E_N} is given by,

$$\Omega^{\mathsf{c}}_{E_N} = \{\omega \in \Omega \;:\; |W_N(\omega)| > 2\epsilon N\}.$$

Since ϵ is arbitrary we can disregard the factor of 2 and square both sides to remove the annoying absolute sign. Thus,

$$\Omega^{\mathsf{c}}_{E_N} = \left\{\omega \in \Omega \;:\; W_N^2(\omega) > N^2\epsilon^2\right\}.$$

Now we can use 7.0.3 in stating,

$$\mu\left(\Omega^{\mathsf{c}}_{E_N}\right) < \frac{1}{N^2\epsilon^2} \int_0^1 W_N^2(\omega).$$

To evaluate the integral we observe,

$$\begin{aligned}
\int_0^1 W_N^2(\omega) &= \int_0^1 \Big(\sum_{k=1}^{N} R_k(\omega)\Big)^2 \\
&= \sum_{k=1}^{N} \int_0^1 R_k^2(\omega) + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \int_0^1 R_i(\omega) R_j(\omega).
\end{aligned}$$

The second summation is 0 by 7.0.2 (1) while the first summation is N because each integral is 1. Thus,

$$\mu\left(\Omega^{\mathsf{c}}_{E_N}\right) \le \frac{1}{N^2\epsilon^2} N = \frac{1}{N\epsilon^2}.$$

Thus,

$$\mu\left(\Omega^{\mathsf{c}}_{E_N}\right) \to 0 \quad \text{as } N \to \infty.$$

□

We are now ready to prove 7.0.1 (2) which is the **SLLN** for the Bernoulli sequence.

Proof. We noted that to show $\mu\,(\Omega^{\mathsf{c}}_E) = 0$, for any ϵ we will have to find simple sets (closed intervals) E_i such that $\Omega^{\mathsf{c}}_E \subset \bigcup\limits_{i=1}^{\infty} E_i$ and $\sum\limits_{i=1}^{\infty} \mu\,(E_i) < \epsilon$. There is $\delta > 0 \in \mathbb{R}$ such that the set,

$$A_n = \{\omega \in \Omega \;:\; |W_n(\omega)| > \delta n\}\,.$$

is not empty for some n. How did we get this set? Consider an $\omega \in \Omega^{\mathsf{c}}_E$. Then $\frac{S_n(\omega)}{n} \not\to \frac{1}{2}$. This is equivalent to saying,

$$\exists\,(\delta)\,\forall\,(N)\,\exists\,(n)\left[n \geq N \,\wedge\, \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| > \delta\right].$$

Since $\left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| = |\frac{W_n(\omega)}{n}|$ the set A_n is not empty.

Moreover, A_n is just a finite union of disjoint h-intervals and is Lebesgue measurable. By 7.0.3 (removing the absolute sign by raising to 4^{th} power)

$$\mu\,(A_n) < \frac{1}{\delta^4 n^4} \int\limits_0^1 (\sum_{k=1}^{n} R_k)^4.$$

The integrand yields 5 kinds of terms,

(1) R_j^4 for $j = 1 \cdots n$.
(2) $R_j^2 R_k^2$ for $j \neq k$.
(3) $R_j^2 R_k R_l$ for $j \neq k \neq l$.
(4) $R_j^3 R_k$ for $j \neq k$.
(5) $R_j R_k R_l R_m$ for $j \neq k \neq l \neq m$.

Using 7.0.2 it is easy to observe that only the first and second kind of terms integrate out to 1 while others give zero. There are n terms of the first kind and $3n(n-1)$ terms involving the second kind. Thus,

$$\mu\,(A_n) < \frac{3}{n^2\delta^4}.$$

The idea is to cover Ω^{c}_E using the sets A_n. Since these sets are not necessarily closed we can take their closure to get closed intervals without changing the measure. Thus it is not necessary to find closed intervals. However, we need to make sure that the sequence of sets decrease in measure so that the countable sum can be made arbitrarily small.

For a constant C, set $\delta_n = Cn^{\frac{-1}{2}}$. Then,

$$\sum_{n=1}^{\infty} \frac{3}{{\delta_n}^4 n^2} = \frac{3}{C} \sum_{n=1}^{\infty} \frac{1}{n^{\frac{3}{2}}}$$

converges and can be made smaller than any epsilon by choosing sufficiently large C.

Let,

$$E_n = \{\omega \in \Omega \ : \ |W_n(\omega)| > \delta_n n\}.$$

If $\Omega_E^{\mathsf{c}} \subset \bigcup_{n=1}^{\infty} E_n$, then

$$\mu\left(\Omega_E^{\mathsf{c}}\right) \leq \sum_{n=1}^{\infty} \mu\left(A_n\right) = \frac{3}{C} \sum_{n=1}^{\infty} \frac{1}{n^{\frac{3}{2}}} < \epsilon.$$

Thus $\mu\left(\Omega_E^{\mathsf{c}}\right) = 0$ provided $\Omega_E^{\mathsf{c}} \subset \bigcup_{n=1}^{\infty} E_n$. However, this is easy to see. Taking complement what we need to show is

$$\bigcap_{n=1}^{\infty} E_n^{\mathsf{c}} \subset \Omega_E.$$

This is easily verified. $\square$

We have seen that remarkable observations can be made by working in a measure theoretic framework. Such arguments are standard in probability study. We isolate a few more concepts that are essential in probability computation. One such technique we already saw when we showed that the set Ω_E is Lebesgue measurable by showing that is countable union and intersection of simple sets that are Lebesgue measurable. In general, to prove that a set is measurable w.r.t a probability measure, we need to show that it is an event in the underlying σ- algebra. This is usually done by showing that that the set is a finite (or countable) combination of countable unions and intersections of elementary sets that are events. We show a couple of examples to highlight this idea before giving a concrete definition.

Example 7.0.5. *Let E be the event where $\sum_{n=1}^{\infty} \frac{R_n(\omega)}{n}$ converges. This is an interesting series. We know that $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges but $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges. Is E a plausible event? In other words is Ω_E Lebesgue measurable? Here Ω_E*

is given by,

$$\left\{\omega \in \Omega \;:\; \sum_{n=1}^{\infty} \frac{R_n(\omega)}{n} < \infty\right\}.$$

Let $T_n(\omega) = \sum\limits_{k=1}^{n} \frac{R_k(\omega)}{k}$. *If* $\omega \in \Omega_E$, *then for any* ϵ, *there is an integer* k *such that* $|T_n(\omega) - T_m(\omega)| < \epsilon$ *whenever* $n, m \geq \epsilon$. *We can get rid of the* ϵ *by finding a positive integer* r *such that* $\frac{1}{r} < \epsilon$. *Consider the set,*

$$A_{m,n,r} = \left\{\omega \in \Omega \;:\; |T_n(\omega) - T_m(\omega)| < \frac{1}{r}\right\}.$$

This set is Lebesgue measurable being a union of finite h-intervals in Ω. *We can write* Ω_E *as,*

$$\Omega_E = \bigcap_{r=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{m,n \geq k}^{\infty} A_{m,n,r}.$$

This shows that Ω_E *is Lebesgue measurable.*

Example 7.0.6. *Let* E *be the event where a prescribed sequence, say* $HTTH$, *occurs infinitely often. To describe* Ω_E, *let* E_n *be the event where the pattern occurs beginning at the* n^{th} *step.* E_n *is described by a finite number of conditions on the Rademacher functions,*

$$\Omega_{E_n} = \{\omega \in \Omega \;:\; R_n(\omega) = 1, R_{n+1}(\omega) = -1, R_{n+2}(\omega) = -1, R_{n+3}(\omega) = 1\}.$$

How do we get Ω_E *from* Ω_{E_n}? *Since the pattern occurs infinitely often, no matter what integer* k *we choose, there should be a* n *far out such that* $\omega \in \Omega_{E_n}$. *Thus,*

$$\Omega_E = \bigcap_{k=1}^{\infty} \bigcup_{n \geq k}^{\infty} \Omega_{E_n}.$$

What can we say about their probabilities? All these examples have a general pattern. Say we have a countable collection of events $\{E_1, E_2, \cdots\}$. Let E be the event that infinitely many of the E_i's occur. Can we determine the probability of E from the probabilities of E_i's? Two famous results—the Borel-Cantelli Lemmas—are relevant to the answer. First we will give a description of the term infinitely often and almost always.

For a sequence of events (E_n),we define the limit superior, and limit inferior as.

$$\limsup E_n = \bigcap_{n\geq 1}\left(\bigcup_{k\geq n} E_k\right)$$
$$\liminf E_n = \bigcup_{n\geq 1}\left(\bigcap_{k\geq n} E_k\right).$$

Remark 7.0.2. *The terminology stems from the corresponding definitions for a sequence of real numbers. For example if we let (s_n) to be a sequence of real numbers then we can make observations of their limiting behavior by looking at the* sup *and* inf *of the tail sets. Let us define the following,*

$$u_n = \inf\{s_k \ : \ k \geq n\}$$
$$v_n = \sup\{s_k \ : \ k \geq n\}$$

Then it is easy to see that $u_1 \leq u_2 \cdots$, while $v_1 \geq v_2 \geq \cdots$. Since this are monotonic sequence, if they converge they will converge to their sup *and* inf *respectively. Thus,*

$$\liminf s_n := \lim_{n\to\infty} u_n = \sup\left\{u_n \ : \ n \in \mathbb{Z}^+\right\}$$
$$\limsup s_n := \lim_{n\to\infty} v_n = \inf\left\{v_n \ : \ n \in \mathbb{Z}^+\right\}$$

Similarly, we can define the same for sequence of sets. However, to get supremum and infimum, we need an order relation. This is done through the relation $\subset$. Thus if we have a sequence of set (E_n), the sup *is defined as $\bigcup\limits_{n=1}^{\infty} E$. This is easy to see; as for any n, $E_n \subset \bigcup\limits_{n=1}^{\infty} E$. Similarly, the* inf *is defined as $\bigcap\limits_{n=1}^{\infty} E$.*

Now analogously we can define,

$$U_n = \inf\{E_k \ : \ k \geq n\}$$
$$V_n = \sup\{E_k \ : \ k \geq n\}$$

Thus, we have $U_1 \subset U_2 \subset U_3 \cdots$*, while* $V_1 \supset V_2 \supset V_3 \cdots$*. Thus we see that* $U_n \nearrow \bigcup_{n=1}^{\infty} U_n$ *while* $V_n \searrow \bigcap_{n=1}^{\infty} V_n$*. Hence,*

$$\liminf E_n := \lim_{n\to\infty} U_n = \sup\left\{U_n \ : \ n \in \mathbb{Z}^+\right\}$$
$$\limsup E_n := \lim_{n\to\infty} V_n = \inf\left\{V_n \ : \ n \in \mathbb{Z}^+\right\}$$

In probability, we denote the $\limsup E_n$ *as the set* $\{E_n \quad i.o\}$ *where the i.o stands for* ***infinitely often****. This is from the fact that if* $\omega \in \limsup E_n$*, then for any* n *there is a* k *such that* $\omega \in E_k$*. Thus no matter how far out we go, we can find a* k *farther such that* $\omega \in E_k$*. Hence,* $\limsup E_n$ *is the event that gives the point in* Ω *which are in infinitely many of the events.*

We denote the set $\liminf E_n$ *as the set* $\{E_n \quad a.a\}$ *where the a.a stands for* ***almost always****. This is from the fact that if* $\omega \in \liminf E_n$*, then there is an* n *such that for any* k *greater than* n*,* $\omega \in E_k$*. Thus,* $\liminf E_n$ *is the event that all but a finite number of events occur.*

If $\limsup E_n = \liminf E_n = E$*, then we say that* $\lim_{n\to\infty} E_n = E$*.*

We will need the following result.

Proposition 7.0.4. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space. If* $\{E_n\} \subset \mathscr{F}$ *is a countable collection of events, then*

$$\mathbb{P}(\liminf E_n) \leq \liminf \mathbb{P}(E_n) \leq \limsup \mathbb{P}(E_n) \leq \mathbb{P}(\limsup E_n).$$

Proof. Note that $(\mathbb{P}(E_n))$ is a sequence of positive numbers and so the middle inequality is trivial. First let us show that,

$$\mathbb{P}(\liminf E_n) \leq \liminf \mathbb{P}(E_n).$$

Consider the set,

$$U_n = \inf\{E_k \ : \ k \geq n\} = \bigcap_{k\geq n} E_k.$$

Then, by 7.0.2, (U_n) is an increasing sequence, and by the properties of a measure (see 3.2.1),

$$\mathbb{P}(\liminf E_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} U_n\right) = \lim_{n\to\infty} \mathbb{P}(U_n).$$

Since $U_n \subset E_k$ for all $k \geq n$, $\mathbb{P}(U_n) \leq \mathbb{P}(E_k)$ for all $k \geq n$. Thus,

$$\mathbb{P}(U_n) \leq \inf\{\mathbb{P}(E_k) \ : \ k \geq n\}.$$

Hence, taking limits

$$\lim_{n\to\infty} \mathbb{P}(U_n) \leq \lim_{n\to\infty} \inf\{\mathbb{P}(E_k) \ : \ k \geq n\} = \liminf \mathbb{P}(E_n).$$

We get the other inequality using an analogous argument. $\square$

Theorem 7.0.2 (Continuity of Probability measure). Let $\{E_n : n \in \mathbb{Z}^+\}$ be a countable collection of events in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. If $E = \lim_{n\to\infty} E_n$, then $\lim_{n\to\infty} \mathbb{P}(E_n) = \mathbb{P}(E)$.

Proof. Since $\lim_{n\to\infty} E_n = \limsup E_n = \liminf E_n$, using 7.0.4 we get the desired result. This theorem highlights the continuity of the probability measure since,

$$\lim_{n\to\infty} \mathbb{P}(E_n) = \mathbb{P}\left(\lim_{n\to\infty} E_n\right),$$

interchanges the limit operation. $\square$

Theorem 7.0.3 (First Borel-Cantelli Lemma). Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a countable collection of events, $\{E_n\} \subset \mathscr{F}$, let $E = \{E_n \text{ i.o}\}$. Then,

$$\text{If } \sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty \quad \text{then } \mathbb{P}(E) = 0.$$

Proof. Let $V_n = \bigcup_{k\geq n} E_k$. Then $E = \bigcap_{n=1}^{\infty} V_n$. Thus $E \subset V_n$ for every n. Now,

$$\mathbb{P}(V_n) \leq \sum_{k=n}^{\infty} \mathbb{P}(E_k).$$

Since, $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$, for any ϵ, there is an N such that $\sum_{k=N}^{\infty} \mathbb{P}(E_k) < \epsilon$. Thus,

$$\mathbb{P}(E) \leq \mathbb{P}(V_N) \leq \sum_{k=N}^{\infty} \mathbb{P}(E_k) < \epsilon.$$

Since ϵ was arbitrary we get the result. $\square$

Example 7.0.7 (Run lengths). *Consider the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where $\Omega = (0, 1]$, $\mathscr{F}$ is the collection of Lebesgue measurable sets in Ω and $\mathbb{P} = \mu$ is the Lebesgue measure. Thus we are in the Borel space of describing infinite coin tosses. For any n define the run length function l_n by,*

$l_n(\omega) = \#$ *consecutive 1's in the binary expansion of ω starting at the n^{th} place.*

Thus $l_n(\omega) = k$ if,

$$R_{n(\omega)} = 1, R_{n+1}(\omega) = 1, \ldots, R_{n+k-1} = 1, R_{n+k} = -1.$$

A run length function gives us the number of consective heads in a sequence ω of infinite coin tosses. Take a sequence of non-negative integers $r_1, r_2, r_3, \ldots$

and let E_n be the event that we have atleast r_n consecutive heads starting at the n^{th} toss. Then,

$$\Omega_{E_n} = \left\{\omega \in \Omega \ : \ R_{n(\omega)} = 1, R_{n+1}(\omega) = 1, \ldots, R_{n+r_n-1} = 1\right\}.$$

Let $E = \{E_n \ i.o\}$. Then,

$$P(E_n) = \mu\left(\Omega_{E_n}\right) = (\frac{1}{2})^{r_n}.$$

By 7.0.3, if $\sum\limits_{n=1}^{\infty} \left(\frac{1}{2}\right)^{r_n} < \infty$, then $P(E) = 0$. As a concrete case take $r_n = n$. Then the probability of the event that consists of n consecutive heads starting at the n^{th} toss, infinitely often, is 0.

Next we define conditional probability and independence of events in a probability space.

Definition 7.0.8 (Conditional probability measure)**.** *Suppose A, B are events in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathbb{P}(B) > 0$. The conditional probability (measure) of A given B is*

$$\mathbb{P}\left(A|B\right) = \frac{\mathbb{P}\left(A \cap B\right)}{\mathbb{P}\left(B\right)}.$$

Remark 7.0.3. *Fix an event $B \in \mathscr{F}$. Then the the conditional probability measure is a function $\mathbb{P}\left(.|B\right) \ : \mathscr{F} \to [0, \infty]$. Recall that $\mathscr{F}_B$ is the restricted sigma algebra. See 3.1.1. If we restrict the sample space to B, the measure space $(B, \mathscr{F}_B, \mathbb{P}\left(.|B\right))$ becomes a probability space. This is proved in the following theorem.*

Theorem 7.0.4 (Conditioned probability space)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For an given $B \in \mathscr{F}$, $(B, \mathscr{F}_B, \mathbb{P}\left(.|B\right))$ is also a probability space.

Proof. Note that $\mathscr{F}_B = \{E \cap B \ : \ E \in \mathscr{F}\}$ is a σ- algebra. Clearly, $\mathbb{P}\left(\emptyset \mid B\right) = 0$. Also,

$$\mathbb{P}\left(B \mid B\right) = \frac{\mathbb{P}\left(B \cap B\right)}{\mathbb{P}\left(B\right)} = 1.$$

Now, let $\{A_i\} \subset \mathscr{F}_B$ be a sequence of pairwise disjoint sets in the restricted sigma algebra $\mathscr{F}_B$.

$$\begin{aligned}
\mathbb{P}\left((\dot{\bigcup}_{i=1}^{\infty} A_i) \;\middle|\; B \right) &= \frac{\mathbb{P}\left((\dot{\bigcup}_{i=1}^{\infty} A_i) \cap B \right)}{\mathbb{P}(B)} \\
&= \frac{\mathbb{P}\left(\dot{\bigcup}_{i=1}^{\infty} (A_i \cap B) \right)}{\mathbb{P}(B)} \\
&= \frac{\sum\limits_{i=1}^{\infty} \mathbb{P}((A_i \cap B))}{\mathbb{P}(B)} \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A_i \mid B)
\end{aligned}$$

□

The conditional probability measure of an event A in a sample space, given the probability measure of an event B in a sample space corresponds to probability of event A occuring knowing that B has occured. We give an example below to highlight this, i.e knowing that B has occured, our sample space is now the restricted probability space $(B, \mathscr{F}_B, \mathbb{P}(.|B))$. Hence calculating the probability of A is calculating the conditional probability measure of A given B.

Example 7.0.8. *Let $\mathcal{B}$ be the probabilistic space of the Bernoulli sequence of infinite coin tosses. Let A be the event that we tossed two consecutive Heads, and B be the event that the first toss was head. What is the probability that A occured given that B occured? Recall that our probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is given by $\Omega = (0, 1]$, $\mathscr{F}$ = Lebesgue measurable sets in Ω, and $\mathbb{P} = \mu$, the Lebesgue measure.*

Since B occured, we need to calulate $P(A|B)$. With A we identify the corresponding set in the sigma algebra as

$$\Omega_A = \{\omega \in \Omega \ : \ R_1(\omega) = 1, R_2(\omega) = 1\}.$$

Hence $P(A) = \mu(\Omega_A) = \mu\left((\frac{3}{4}, 1]\right) = \frac{1}{4}$. Similary $P(B) = \mu(\Omega_B) = \mu\left((\frac{1}{2}, 1]\right) = \frac{1}{2}$. Thus

$$P(A|B) = \mu(\Omega_A|\Omega_B) = \frac{\mu(\Omega_A \cap \Omega_B)}{\mu(\Omega_B)} = \frac{1}{2}.$$

The following theorem is very important whose proof is almost trivial.

Theorem 7.0.5 (Total probability theorem)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. If $\{H_i\}$ is a countable collection of pairwise disjoint events such that $\mathbb{P}(H_i) \neq 0$ for all i and $\bigcup_{i=1}^{\infty} H_i = \Omega$, then for any event $A \in \mathscr{F}$,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \mid H_i)\, \mathbb{P}(H_i).$$

Proof. Since $A \subset \Omega$, we can observe that $A \subset \dot{\bigcup}_{i=1}^{\infty} H_i$. Thus $A = \dot{\bigcup}_{i=1}^{\infty} (A \cap H_i)$. Hence,

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}\left(\dot{\bigcup}_{i=1}^{\infty} (A \cap H_i)\right) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A \cap H_i) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(A \mid H_i)\, \mathbb{P}(H_i).
\end{aligned}$$

$\square$

If knowing that B occured doesn't change the probability of A, then we know that A and B are independent. For example if B is event that the first coin toss is a Head and A is the event that the second toss also a Head, then A is independent of B. That is $P(A|B) = P(A)$. We make this notion precise w.r.t a probability space.

Definition 7.0.9 (Independence of events)**.** *Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, two events A, B are independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B).$$

Note that in 7.0.8, the events A, B are NOT independent. Also, observe that the definition is symmetric.

Proposition 7.0.5. *If A_1, A_2 are independent events in a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, then so are A_1^{c} and A_2.*

Proof. Note that, $A_2 = (A_2 \cap A_1) \dot{\bigcup} (A_2 \cap A_1^{\mathsf{c}})$. Thus,

$$\begin{aligned}
\mathbb{P}(A_2) &= \mathbb{P}(A_2 \cap A_1) + \mathbb{P}(A_2 \cap A_1^{\mathsf{c}}) \\
&\implies \mathbb{P}(A_2) - \mathbb{P}(A_2 \cap A_1) = \mathbb{P}(A_2 \cap A_1^{\mathsf{c}}) \\
&\implies \mathbb{P}(A_2) - \mathbb{P}(A_2)\, \mathbb{P}(A_1) = \mathbb{P}(A_2 \cap A_1^{\mathsf{c}}) \\
&\implies \mathbb{P}(A_2)\, (1 - \mathbb{P}(A_1)) = \mathbb{P}(A_2 \cap A_1^{\mathsf{c}}) \\
&\implies \mathbb{P}(A_2)\, \mathbb{P}(A_1^{\mathsf{c}}) = \mathbb{P}(A_2 \cap A_1^{\mathsf{c}})
\end{aligned}$$

□

We extend this idea for a collection of events.

Definition 7.0.10 (Independence of a collection of events and sigma algebras)**.** *A collection of events $\{A_i\}$ are independent if for each $n \in \mathbb{Z}^+$ and each selection of integers, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,*

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j}).$$

We say a collection of σ- algebras $\{\mathscr{F}_i\}$ are independent if each collection of events chosen individually from the σ- algebras are independent.

Example 7.0.9. *Three events A, B, C in the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ are independent if*

(1) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
(2) $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$.
(3) $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$.
(4) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Example 7.0.10. *Let H_i be the event that the i^{th} toss is a head in the space of Bernoulli sequences $\mathcal{B}$. Easy to check that the corresponding events $\Omega_{H_i} = \{\omega \in \Omega \ : \ R_i(\omega) = 1\}$ are independent.*

Proposition 7.0.6. *If $\{A_i\}$ are independent then so are $\{A_i^{\mathsf{c}}\}$.*

Theorem 7.0.6 (Second Borel-Cantelli Lemma)**.** Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\{A_i\}$ be a collection of independent events. Let $A = \{A_n \text{ i.o}\}$.

$$\text{If} \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty \quad \text{then} \, \mathbb{P}(A) = 1.$$

Proof. $A = \limsup A_k = \bigcap_{k \geq 1} \left(\bigcup_{n \geq k} A_n \right)$ and so $A^{\mathsf{c}} = \bigcup_{k=1} \bigcap_{n \geq k} A_n^{\mathsf{c}}$. To show that $\mathbb{P}(A) = 1$, we will show that $\mathbb{P}(A^{\mathsf{c}}) = 0$. It suffices to show that $\mathbb{P}\left(\bigcap_{n \geq k} A_n^{\mathsf{c}} \right) = 0$ for all k. Due to independence,

$$\mathbb{P}\left(\bigcap_{n \geq k}^{l} A_n^{\mathsf{c}} \right) = \prod_{n=k}^{l} \mathbb{P}(A_n^{\mathsf{c}}).$$

Now, $\mathbb{P}\left(A_n^{\mathsf{c}}\right) = 1 - \mathbb{P}\left(A_n\right)$. Since $1 - x \leq e^{-x}$, $\mathbb{P}\left(A_n^{\mathsf{c}}\right) = 1 - \mathbb{P}\left(A_n\right) \leq e^{-\mathbb{P}(A_n)}$. Hence,

$$\mathbb{P}\left(\bigcap_{n\geq k}^{l} A_n^{\mathsf{c}}\right) \leq \prod_{n=k}^{l} e^{-\mathbb{P}(A_n)} = e^{-\sum\limits_{n=k}^{l}\mathbb{P}(A_n)}.$$

Thus, as $l \to \infty$, $\mathbb{P}\left(A_n^{\mathsf{c}}\right) \to 0$. □

Example 7.0.11. *Let H_n be the event of a head at the n^{th} toss. The corresponding event in the σ- algebra is*

$$\Omega_{H_n} = \left\{\omega \in \Omega \ : \ R_n(\omega) = 1\right\}.$$

Furthermore, $\mathbb{P}\left(\Omega_{H_n}\right) = \frac{1}{2}$. Since $\sum\limits_{i=1}^{\infty} \frac{1}{2} = \infty$, the probability that we see Heads infinitely often is 0.

Example 7.0.12. *Any finite pattern in an infinite sequence of coin tosses occurs infinitely often with probability of* 1*. To see a concrete case, consider the sequence $HTTH$. Let E_n be the event where $HHTH$ occur starting at step n and let Ω_{E_n} be the corresponding set in the σ- algebra. Then*

$$\Omega_{E_n} = \Omega_{H_n} \cap \Omega_{H_{n+1}}^{\mathsf{c}} \cap \Omega_{H_{n+2}}^{\mathsf{c}} \cap \Omega_{H_{n+3}},$$

where Ω_{H_n} is described in 7.0.11. From independence $\mathbb{P}\left(\Omega_{E_n}\right) = \frac{1}{2^4}$. Since Ω_{E_n} and $\Omega_{E_{n+1}}$ are not independent we cannot apply 7.0.6. However, $\{\Omega_{E_n}, \Omega_{E_{n+4}}, \Omega_{E_{n+8}}, \cdots\}$ are independent that satisy,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\Omega_{E_{4n+1}}\right) = \infty.$$

So, $\mathbb{P}\left(\{\Omega_{E_{4n+1}}\ i.o\}\right) = 1$. But,

$$\{\Omega_{E_{4n+1}}\ i.o\} \subset \{\Omega_{E_n}\ i.o\}.$$

Hence, $\mathbb{P}\left(\Omega_{E_n}\ i.o\}\right) = 1$, since probability measure of any event is bounded by 1*.*

From now onwards, we will not be explicit in stating an underlying probabilistic process. We will denote the probability of an event by its probability measure.

Chapter 8

Random Variables and Expectation

8.1. Random variables and their distributions

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $(\mathbb{Y}, \mathscr{N})$ be a measure space. A $(\mathscr{F}, \mathscr{N})$ – measurable function X is called a $\mathbb{Y}$ valued random variable. In particular for $\mathbb{Y} = \mathbb{R}$ or $\mathbb{Y} = \mathbb{R}^n$ we have the following definition,

Definition 8.1.1 (Random variable)**.** *A random variable (r.v) X is a real valued $(\mathscr{F}, \mathfrak{B}_{\mathbb{R}})$ – measurable measurable function that maps from $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$. A random vector X is a vector valued $(\mathscr{F}, \mathfrak{B}_{\mathbb{R}^n})$ – measurable function that maps from $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$.*

As always, we wish to discard the importance of behavior of sets of measure zero.

Definition 8.1.2 (Almost surely)**.** *Suppose X, Y are two random variables on the same probability space. Then $X = Y$ a.s. or $X = Y$ almost surely, means that, $\mathbb{P}(\{\omega \in \Omega \ : \ X(\omega) \neq Y(\omega)\}) = 0$.*

Note, almost surely is precisely what is meant by almost everywhere that was defined for a general measure space.

Example 8.1.1. *Consider the space $\mathcal{B}$ of Bernoulli sequence and the corresponding measure space $\Omega = (0, 1]$ with the Lebesgue measure μ. Any Rademacher function R_n which maps from $((0, 1], \mathscr{L}, \mu)$ into $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ is a random variable because for any $E \in \mathfrak{B}_{\mathbb{R}}$, $R_n^{-1}(E) = \{\omega \in (0, 1] \ : \ R_n(\omega) \in E\}$ is a finite union of intervals. Similarly W_N is a random variable.*

Example 8.1.2. *Consider the space $\mathcal{B}$ of Bernoulli sequence and let X be a function from $\mathcal{B} - \mathcal{B}_{neg}$ in to $(0, 1]$ defined by,*

$$X(b_1, b_2, b_3, \dots) = 0.\omega_1\omega_2 \dots,$$

where $\omega_i = 1$ if b_i is a head H, $\omega_i = 0$ if b_i is a tails T. Then X is a random variable. Thus a random variable transforms *a probabilistic process to a measure space.*

Proposition 8.1.1. *Let X be a r.v. from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to a measure space $(\mathbb{Y}, \mathscr{N})$. Let $\mathbb{P}_X$ be the induced measure in $(\mathbb{Y}, \mathscr{N})$. Then, $(\mathbb{Y}, \mathscr{N}, \mathbb{P}_X)$ is a probability space.*

Proof. By 4.1.3, we know that $(\mathbb{Y}, \mathscr{N}, \mathbb{P}_X)$ is a measure space. Hence, we only need to check if $\mathbb{P}_X(\mathbb{Y}) = 1$. Indeed $X^{-1}(\mathbb{Y}) = \Omega$ and $\mathbb{P}(\Omega) = 1$, and thus we see that $(\mathbb{Y}, \mathscr{N}, \mathbb{P}_X)$ is a probability space. $\square$

Definition 8.1.3 (Distribution)**.** *The probability measure $\mathbb{P}_X$ on $(\mathbb{Y}, \mathscr{N})$ is called the distribution of the r.v. X and is said to be induced by X.*

Theorem 8.1.1. Suppose X, Y are two r.v. mapping a (complete) probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to a measurable space $(\mathbb{Y}, \mathscr{N})$, such that $X = Y$ a.s. Then X, Y have the same distribution.

Proof. Let $N = \{\omega \in \Omega \ : \ X(\omega) \neq Y(\omega)\}$. Then, by our assumption, $\mathbb{P}(N) = 0$. Let $B \in \mathscr{N}$ be arbitrary. Then

$$X^{-1}(B) = (X^{-1}(B) \cap N) \dot{\bigcup} (X^{-1}(B) \cap N^{\mathsf{c}}).$$

Thus,

$$\begin{aligned}\mathbb{P}_X(B) &= \mathbb{P}\left(X^{-1}(B)\right) \\ &= \mathbb{P}\left(X^{-1}(B) \cap N\right) + \mathbb{P}\left(X^{-1}(B) \cap N^{\mathsf{c}}\right) \\ &\leq 0 + \mathbb{P}\left(Y^{-1}(B)\right) \\ &= \mathbb{P}_Y(B)\end{aligned}$$

Similarly, $\mathbb{P}_Y(B) \leq \mathbb{P}_X(B)$. $\square$

Even when we don't have a complete probability space, we can easily complete it by 3.3.1. Hence we can easily forget if the probability space is complete or not. We have seen that there are two notions of equality for a random variable.

Definition 8.1.4. *Two r.v. s X, Y are equal in distribution if for all $B \in \mathscr{N}$,*

$$\mathbb{P}\{X \in B\} = \mathbb{P}\{Y \in B\}.$$

Note that the set $\{X \in B\}$ is shorthand for the set $\{\omega \in \Omega \ : \ X(\omega) \in B\}$.

Definition 8.1.5. *Two r.v. s X, Y are equal a.s. if*

$$\mathbb{P}\left(\{\omega \ : \ X(\omega) = Y(\omega)\}\right) = 1.$$

By 8.1.1, we know that if two random variables are equal almost surely, then they are equal in distribution. However the converse is not true as the following example illustrates.

Example 8.1.3. *Toss a fair coin and set X to be 1 if the outcome is heads and 0 otherwise, and let Y be 0 if the outcome is heads and 1 otherwise. Then, $\mathbb{P}\, X = 1 = \mathbb{P}\, X = 0 = \frac{1}{2}$. Similarly, $\mathbb{P}\, Y = 1 = \mathbb{P}\, Y = 0 = \frac{1}{2}$. So that X and Y are equal in distribution but are not equal almost surely.*

Given a r.v. induces probability measure on its co-domain, we can define new r.v on the range. We have the following relationship,

Proposition 8.1.2. *Let X be $\mathbb{Y}$ valued r.v. from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to a measurable space $(\mathbb{Y}, \mathscr{N})$. Let Y be a $(\mathscr{N}, \mathscr{O}) -$ measurable function from $(\mathbb{Y}, \mathscr{N})$ in to $(\mathbb{W}, \mathscr{O})$. Let $\mathbb{P}_X$ be the distribution on $(\mathbb{Y}, \mathscr{N})$. Then Y is a $\mathbb{W}$ valued r.v. on $(\mathbb{Y}, \mathscr{N}, \mathbb{P}_X)$ with the same distribution as the $\mathbb{W}$ valued r.v. $Y \circ X$ on $(\Omega, \mathscr{F}, \mathbb{P})$.*

Proof. It follows the same ideas as in 4.1.2. □

Since a r.v. is a measurable function, all the properties in 4.2.1-4.2.5 apply to a r.v. We now specialize to the case of random variables mapping into $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$. From now on, each r.v. is a measurable mapping in to the reals (or extended reals). Thus we fix the following notations that will be used throughout.

$$\text{Probability space: } (\Omega, \mathscr{F}, \mathbb{P}),$$
$$\text{Random variable } X{:}\mathscr{F} - \text{measurable mapping.}$$

Clearly for every Borel set $B \in \mathfrak{B}_{\mathbb{R}}$, $X^{-1}(B) \in \mathscr{F}$. What is the smallest sigma-algebra on Ω on which X is measurable? If we just collect the iverse image of X in $\mathscr{F}$, the collection is the pre-image sigma algebra 3.1.4 and by definition X is measurable on that sigma algebra.

Definition 8.1.6. *If X is a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$, then σX is the smallest sigma-algebra on Ω for which X is σX-measurable.*

By the remarks preceeding the definition, it is not hard to see that

Proposition 8.1.3. *σX contains sets of the form $X^{-1}(B)$ for some $B \in \mathfrak{B}_{\mathbb{R}}$.*

Thus, $\sigma X = X^{-1}(\mathfrak{B}_{\mathbb{R}})$.

Remark 8.1.1. *Given a collection of random variables on* $(\Omega, \mathscr{F}, \mathbb{P})$, $\mathscr{C} = \{X_\alpha : \alpha \in A, X_\alpha r.v\}$, *what is the smallest sigma algebra that makes each* X_α *measurable on it. From our discussions on product sigma algebra it is not hard to see that*

$$\sigma\left(\bigcup_{\alpha\in A} \sigma(X_\alpha)\right),$$

is the smallest sigma-algebra on which each X_α *is measurable. This, along with the fact that,* $\mathfrak{B}_{\mathbb{R}^n} = \bigotimes_{1=1}^{n} \mathfrak{B}_{\mathbb{R}}$, *gives us the following result*

Proposition 8.1.4. *Let* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ *be a random vector on* $(\Omega, \mathscr{F}, \mathbb{P})$. *Then,*

$$\sigma\boldsymbol{X} = \sigma\left(\bigcup_{i=1}^{n} \sigma(X_i)\right).$$

Let us define a distribution function for the case of a probability space.

Definition 8.1.7 (Distribution function)**.** *A real valued function* F *defined on* $\mathbb{R}$ *is a distribution function if it is monotone increasing, right continuous, and*

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1.$$

Given a probability measure $\mathbb{P}$ on $\mathbb{R}$, we can always define a distribution function $F : \mathbb{R} \to [0,1]$ as $F(x) = \mathbb{P}((-\infty, x])$ and vice versa by 3.5.2. Thus if we have a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a r.v. on it, we get an induced probability measure (distribution) $\mathbb{P}_X$ on $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, which then corresponds to a distribution function $F_X : \mathbb{R} \to [0,1]$ and is given by $F_X(x) = \mathbb{P}_X([-\infty, x])$. Let us make these ideas precise

Definition 8.1.8. *Let* X *be a r.v. on* $(\Omega, \mathscr{F}, \mathbb{P})$. *The distribution function (d.f.) of* X, *denoted by* F_X, *is given by*

$$F(x) = \mathbb{P}(X \leq x),$$

for each $x \in \mathbb{R}$.

Theorem 8.1.2. Let X be a r.v. and let F_X be its d.f. Then,

(1) F_X is non-decreasing, right continuous, real valued function with left hand limits given by,

$$F_X(x-) = \lim_{x_n \nearrow x} F(x_n).$$

(2)

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1.$$

(3) F_X has at most a countable number of discontinuities.

Thus,

$$X \rightsquigarrow F_X.$$

Is the converse true, i.e.

$$F \rightsquigarrow X_F?$$

That is, given a distribution function on $\mathbb{R}$, is there a r.v. X on some probability space such that $F_X = F$? The following theorem due to Skohord answers this question.

Theorem 8.1.3. If a function $F : \mathbb{R} \to [0,1]$ is a distribution function, then there exists a r.v. $X : [0,1] \to \mathbb{R}$ defined on the probability space $\left([0,1], \mathscr{L}_{[0,1]}, \mu\right)$ such that $F = F_X$, where $\mathscr{L}_{[0,1]}$ is the family of Lebesgue measurable sets on $[0,1]$ and μ is the corresponding Lebesgue measure.

Proof. Let us define $X : [0,1] \to \mathbb{R}$ by,

$$X(\omega) = \inf \{x \in \mathbb{R} \ : \ F(x) \geq \omega\}, \quad 0 \leq \omega \leq 1.$$

First, we will show that X is a $\left(\mathscr{L}_{[0,1]}, \mathfrak{B}_{\mathbb{R}}\right)-$measurable function and hence a r.v. Fix an $a \in \mathbb{R}$ and consider the set $\{X \leq a\}$. Since F is increasing this set is just an interval $[0,c]$ where c is the $\sup \{\omega \ : \ X(\omega) \leq a\}$. Hence X is a r.v.

To show that $F = F_X$ we need to show that for any $y \in \mathbb{R}$,

$$\begin{aligned} F(y) = F_X(y) &= \mu_X\left((-\infty, y]\right), \\ &= \mu\left(X^{-1}\left((-\infty, y]\right)\right), \\ &= \mu\left(\{\omega \in \Omega \ : \ X(\omega) \leq y\}\right). \end{aligned}$$

Let $A = \{\omega \ : \ X(\omega) \leq y\}$. Then $A = [0,c]$ where $c = \sup \{\omega \ : \ X(\omega) \leq y\}$. Hence $\mu(A) = c$. Thus we need to show that $F(y) = c = \sup \{\omega \ : \ X(\omega) \leq y\}$. First note that if $\omega = F(y)$ then $X(\omega) = y$ because F is increasing. Thus $F(y) \in A$. If we show that $F(y)$ is an upper bound for A, then we are done. If $\omega \in A$, then $X(\omega) \leq y$. Thus $F(X(\omega)) \leq F(y)$. Since F is right continuous at $X(\omega)$, for any ϵ there is a δ such that $F(X(\omega) + \delta) - F(X(\omega)) < \epsilon$. Now $X(\omega)$ is the infimum of all those $x \in \mathbb{R}$ such that $F(x) \geq \omega$. Hence by definition of infimum, there is an $x_0 \in \mathbb{R}$ such that $x_0 < X(\omega) + \delta$ and $F(x_0) \geq \omega$. Thus, because F is increasing,

$$F(X(\omega) + \delta) \geq F(x_0) \geq \omega.$$

Hence $\omega < \epsilon + F(X(\omega))$. Since ϵ was arbitrary we get $\omega \leq F(X(\omega))$. Hence $F(y)$ is an upper bound for A. $\square$

8.2. Expectations and moments

Appendix A

Preliminary Concepts: Set theory

A.1. Foundations of set theory

In this section, we will cover the basics of axiomatic set theory and will conclude with the Axiom of Choice. Most proofs will be omitted. The axiomatic theory of set is just based two (undefined) notions *class* and the *membership relation* denoted by $\in$. All objects are classes. However there are two kinds of classes:

- Sets
- Proper Classes

If x, A are classes then the expression $x \in A$ means that x is an element of A. This leads to our first definition

Definition A.1.1. *Let x be a class. If x is belongs to some class A then x is called an element.*

All elements are denoted by lower case letters. Hence whenever we write x, y, z we mean classes that belong to some class. Whenever we denote classes by capital letters A, B, C then such a class may be an element of some other class or may not be an element at all.

Definition A.1.2. *Let A, B be classes. We define $A = B$ to mean that every class that has A as its element must have B as an element and vice versa. Logically,*

$$A = B \textit{ iff } (\forall X) \left[A \in X \implies B \in X \wedge B \in X \implies A \in X \right].$$

The first axiom is called the **axiom of Extent** and is an equivalent statement of the above definition.

Axiom 1: $A = B$ iff $x \in A \iff x \in B$.

Definition A.1.3. *Let A, B be classes; we define $A \subseteq B$ to mean that every element of A is an element of B. A is called a subclass of B.*

The second axiom is called the **axiom of class construction** and defines way to construct sets from elements. For this we define a *property* as,

Definition A.1.4. *A property $P(x)$ is a mathematical statement involving an element x such that it can be expressed entirely in terms of the logical symbols $\in, \wedge, \vee, \neg, \exists, \forall$ and variables $x, y, z, A, B \ldots$.*

Axiom 2: If $P(x)$ is a property then there exists a class C whose elements are precisely those that satisfy $P(x)$. Logicall we denote C as,

$$C = \{x : P(x)\} .$$

If A and B are classes then the following properties gives very important classes,

(1) $P(x)$ is $x \in A \vee x \in B$.
(2) $P(x)$ is $x \in A \wedge x \in B$.

The class satisfying the first property is called the union of A, B and is denoted by $A \cup B$. The class satisfying the second property is called the intersection of A, B and is denoted by $A \cap B$.

Definition A.1.5. *The universal class $\mathcal{U}$ is the class of all the elements. Thus it contains classes that belong to some class. Thus,*

$$\mathcal{U} = \{x : x = x\}$$

Definition A.1.6. *The empty class $\emptyset$ is the class that has no elements. Thus,*

$$\emptyset = \{x : x \neq x\}$$

Definition A.1.7. *If two classes A, B have no elements in common, they are said to be disjoint. Thus, A, B are disjoint if,*

$$A \cap B = \emptyset.$$

Definition A.1.8. *The complement of a class A, A^{c} is the class of all elements that do not belong to A. Thus,*

$$A^{\mathsf{c}} = \{x : x \notin A\}$$

Note that for any class A, $\emptyset \subseteq A$, $A \subseteq \mathcal{U}$ and $A \cup A^{\mathsf{c}} = \mathcal{U}$ and $A \cap A^{\mathsf{c}} = \emptyset$. The Demorgan Laws provide a duality about union and intersection,

- $(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}$.
- $(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$.

It is very important fact that union, intersection and complement describe an algebra of classes which is summarized below (easily proven),

- Identity Laws:
  - $A \cup A = A$.
  - $A \cap A = A$.
- Associative Laws:
  - $A \cup (B \cup C) = (A \cup B) \cup C$.
  - $A \cap (B \cap C) = (A \cap B) \cap C$.
- Commutative Laws:
  - $A \cup B = B \cup A$.
  - $A \cap B = B \cap A$.
- Distributive Laws:
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Definition A.1.9. *The difference of two classes A, B, $A - B$ is the class of those elements that are elements of A but not of B. Thus,*

$$A - B = \{x : x \in A \wedge x \notin B\}$$

Note that $A - B = A \cap B^{\mathsf{c}}$.

If a is an element then from Axiom 2 we can construct the class that contains only a. Such a class is called a *singleton* and is given by:

$$\{a\} = \{x : x = a\}.$$

Similarly we can create the *un-ordered* pair $\{a, b\}$ which is,

$$\{a, b\} = \{x : x = a \vee x = b\}.$$

It is easy to see that $\{a, b\} = \{c, d\}$ when $(a = c) \vee (b = d)$ or $(a = d) \vee (b = c)$. Something that is very important is the notion of ordered pair which we denote by (a, b). What is important about ordered pairs is that if $(a, b) = (c, d)$ the $a = c$ and $b = d$. Thus the *order* in which they appear in the set is important.

Definition A.1.10. *If a, b are elements, then the ordered pair is the class given by,*

$$(a, b) = \{\{a\}, \{a, b\}\}$$

Definition A.1.11. *The Cartesian product of two classes A and B denoted by $A \times B$ is the class of all ordered pairs (x, y) where $x \in A$ and $y \in B$. Thus,*

$$A \times B = \{(x, y) : x \in A \wedge y \in B\}.$$

A class of ordered pairs is called a *Graph*. Thus any subclass of $\mathcal{U} \times \mathcal{U}$ is a graph. If G is a graph, we denote G^{-1} to be the inverse graph given by,

$$G^{-1} = \{(y, x) : (x, y) \in G\}.$$

Definition A.1.12. *If G, H are graphs, then $G \circ H$ is the graph defined as follows:*

$$G \circ H = \{(x, y) \ : \ \exists z \ni (x, z) \in H \ \wedge \ (z, y) \in G\}$$

Theorem A.1.1. If G, H, J are graphs, then the following hold:

(1) $(G \circ H) \circ J = G \circ (H \circ J)$.
(2) $\left(G^{-1}\right)^{-1} = G$.
(3) $(G \circ H)^{-1} = H^{-1} \circ G^{-1}$.

Proof. We prove in order,

(1) let $(x, y) \in (G \circ H) \circ J$. Then there is a $z \ni (x, z) \in J$ and $(z, y) \in (G \circ H)$. Thus there is a $u \ni (z, u) \in H$ and $(u, y) \in G$. Thus $(x, u) \in H \circ J$. And so $(x, y) \in G \circ (H \circ J)$. The argument can be reversed and so we get the equality of classes.

(2) Let $(x, y) \in \left(G^{-1}\right)^{-1}$. Hence $(y, x) \in G^{-1}$. Thus $(x, y) \in G$. The other direction is similar.

(3) Let $(x, y) \in (G \circ H)^{-1}$. Hence $(y, x) \in (G \circ H)$. Thus there is a $z \ni (y, z) \in H$ and $(z, x) \in G$. Thus $(z, y) \in H^{-1}$ and $(x, z) \in G^{-1}$, which is just $(x, z) \in G^{-1}$ and $(z, y) \in H^{-1}$. Hence $(x, y) \in H^{-1} \circ G^{-1}$.

□

Definition A.1.13. *Let G be a graph. By the domain of G we mean the class*

$$dom\, G = \{x \ : \ \exists y \ni (x, y) \in G\}.$$

Definition A.1.14. *Let G be a graph. By the range of G we mean the class*

$$range\, G = \{y \ : \ \exists x \ni (x, y) \in G\}.$$

Theorem A.1.2. IF G, H are graphs then

(1) $dom\, G = range\, G^{-1}$.
(2) $dom\, G^{-1} = range\, G$.
(3) $dom\, (G \circ H) \subseteq dom\, H$.
(4) $range\, (G \circ H) \subseteq range\, G$.

The proof of the third statement is as follows,

Proof. Let $x \in dom\, (G \circ H)$. Thus there is a $y \ni (x, y) \in G \circ H$. Thus there is a $z \ni (x, z) \in H$ and $(z, y) \in G$. Thus the existence of $z \ni (x, z) \in H$ means that $x \in dom\, H$. □

An important corollary of the above theorem is that if $range\, H = dom\, G$ then $dom\, G \circ H = dom\, H$. To prove the equality we just need to show that $dom\, H \subseteq dom\, G \circ H$. Consider an element $x \in dom\, H$. Thus there is a $z \ni (x, z) \in H$. Since range of H equal to $dom\, G$, this means that $z \in dom\, G$. Hence there is a $y \ni (z, y) \in G$ and so $(x, y) \in G \circ H$. Thus $x \in dom\, G \circ H$.

Definition A.1.15. *An indexed class is the class denoted by* $\{A_i : i \in I\}$ *where* I *is the class whose elements are called indices.*

Formally an indexed class is a graph G and each $A_i = \{x : (i, x) \in G\}$. Thus if $I = \{1, 2\}$ and $A_1 = \{a, b\}$ and $A_2 = \{e, f\}$ then the indexed class $\{A_i : i \in I\}$ is the graph $G = \{(1, a), (1, b), (2, e), (2, f)\}$.

Definition A.1.16. *Let* $\{A_i : i \in I\}$ *be an indexed family. Then,*

(1) The union of the classes A_i *consists of all those elements* x *that are contained in atleast one* A_i.

$$\bigcup_{i \in I} A_i = \{x \ : \ \exists j \ni x \in A_j\} .$$

(2) The intersection of the classes A_i *consists of all those elements* x *that are contained in each* A_i.

$$\bigcap_{i \in I} A_i = \{x \ : \ \forall j, x \in A_j\} .$$

Theorem A.1.3. Let $\{A_i : i \in I\}$ be an indexed class and B be any class. Then,

(1) If $B \subseteq A_i$ for every $i \in I$ then $B \subseteq \bigcap_{i \in I} A_i$.
(2) If $A_i \subseteq B$ for every $i \in I$ then $\bigcup_{i \in I} A_i \subseteq B$.

The Generalized DeMorgan's Laws and Distributive laws can be restated as:

Theorem A.1.4 (DeMorgan's Laws). Let $\{A_i : i \in I\}$ be an indexed class. Then,

(1) $(\bigcup_{i\in I} A_i)^{\mathsf{c}} = \bigcap_{i\in I} {A_i}^{\mathsf{c}}$
(2) $(\bigcap_{i\in I} A_i)^{\mathsf{c}} = \bigcup_{i\in I} {A_i}^{\mathsf{c}}$

Theorem A.1.5 (Distributive Laws). Let $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$ be indexed classes. Then,

(1) $(\bigcup_{i\in I} A_i) \bigcap (\bigcup_{j\in J} B_j) = \bigcup_{(i,j)\in I\times J} A_i \cap B_j$.
(2) $(\bigcap_{i\in I} A_i) \bigcup (\bigcap_{j\in J} B_j) = \bigcap_{(i,j)\in I\times J} A_i \cup B_j$.

Theorem A.1.6. Let $\{G_i : i \in I\}$ be a family of graphs. Then,

(1) $dom\,(\bigcup_{i\in I} G_i) = \bigcup_{i\in I}(dom\, G_i)$.
(2) $range\,(\bigcup_{i\in I} G_i) = \bigcup_{i\in I}(range\, G_i)$.

In the begining we noted that there were two kinds of classes. We now define the most important kind of class called *Set*.

Definition A.1.17. *A class X is called a set if there is a class Y such that $X \in Y$.*

If for all class Y, $X \notin Y$ then X is called a proper class. The remaining axioms all concern sets.

Axiom 3: Every subclass of a set is itself a set.

Such a subclass is called a *subset*. Note that for any class B, if A is a set then $A \cap B \subseteq A$. Thus *intersections* are sets. The next axiom gives the existence of sets.

Axiom 4: The empty class $\emptyset$ is a set.

Axiom 5: If a, b are sets then the un-ordered pair $\{a, b\}$ is a set.

Note that $\emptyset$ is a set and the set containing the emptyset $\{\emptyset\}$ is also a set i.e we used $b = a$ in the above axiom to construct this set. Thus we can form a new set $\{\emptyset, \{\emptyset\}\}$. We can continue this forever.

Definition A.1.18. *Let A be a set. By the power set of A we mean the class which contains all subsets of A. Thus,*

$$\mathscr{P}(A) = \{X \ : \ X \subseteq A\}.$$

The next two axioms concerns sets of sets.

Axiom 6: If $\mathcal{A}$ is a set of sets then $\bigcup \mathcal{A}$ is also a set.

Note that $\bigcup \mathcal{A}$ is the set $\{x \ : \ \exists A \in \mathcal{A} \ni x \in A\}$.

Axiom 7: If A is a set then $\mathscr{P}(A)$ is also a set.

The following theorem shows that the cartesian product of two sets is also a set.

Theorem A.1.7. If A, B are sets then $A \times B$ is also a set.

Proof. We will show that $A \times B$ is an subset of $\mathscr{P}(\mathscr{P}(A \cup B))$ and thus by Axiom 3 is a set. Let $(x, y) \in A \times B$. Note that $(x, y) = \{\{x\}, \{x, y\}\}$. But $\{x\} \in \mathscr{P}(A \cup B)$ and $\{y\} \in \mathscr{P}(A \cup B)$. Thus $\{\{x\}, \{x, y\}\}$ is a subset of $\mathscr{P}(A \cup B)$. That is $(x, y) \in \mathscr{P}(\mathscr{P}(A \cup B))$. Hence $A \times B$ is a subset of $\mathscr{P}(\mathscr{P}(A \cup B))$. $\square$

Now we will look into the set-theoretic definition of functions. A function f is a *triple* (A, B, f) where A, B are sets and $f \subseteq A \times B$ is a graph satisfying the following conditions:

F 1: For every $x \in A$ there is a $y \in B$ such that $(x, y) \in f$.

F 2: For every $x \in A$, if $y_1, y_2 \in B$ such that $(x, y_1) \in f$ and $(x, y_2) \in f$ then $y_1 = y_2$.

We usually denote (A, B, f) as $f : A \to B$ and write $(x, y) \in f$ as $f(x) = y$. Thus the above conditions become:

F 1: For every $x \in A$ there is a $y \in B$ such that $f(x) = y$.

F 2: For every $x \in A$, $f(x) = y_1$ and $f(x) = y_2$ implies $y_1 = y_2$.

We note that if A, B are sets then any graph $f \subseteq A \times B$ is a function iff

(1) $\mathit{dom}\, f = A$.
(2) $\mathit{range}\, f \subseteq B$.
(3) $F2$ is satisfied.

Some important functions are:

(1) INJECTIVE, A function $f : A \to B$ is said to be injective iff for $x_1, x_2 \in A$, $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.
(2) SURJECTIVE, A function $f : A \to B$ is surjective iff for every $y \in B$ there is a $x \in A$ such that $y = f(x)$, i.e range $f = B$.
(3) BIJECTIVE, A function $f : A \to B$ is bijective iff it is both surjective and injective.

Some examples are as follows:

(1) Identity function. The function $i_A : A \to A$ given by $i_A(x) = x$ for every $x \in A$ is called the identity function.
(2) Inclusion function. Let A, B be sets such that $B \subseteq A$. The function $i_B : B \to A$ is called the inclusion function. Note that $i_B(x) = x$ for every $x \in B$.

(3) Characteristic function. Let 2 designate the class of all functions with two elements, say the class $\{0, 1\}$. If $B \subset A$, then the characteristic function of B is given by $\chi_B : B \to 2$ such that whenever $x \in B$ then $\chi_B(x) = 1$, otherwise $\chi_B(x) = 0$.

(4) Restriction function. Let $C \subseteq A$ and $f : A \to B$. Then the restriction of f to C is the function given by $f_{|C} : C \to B$ such that $f_{|C}(x) = f(x)$ for every $x \in C$.

The next theorem concerns composition.

Theorem A.1.8. If $f : A \to B$ and $g : B \to C$ are functions then $g \circ f : A \to C$ is a function and $(g \circ f)(x) = g(f(x))$ for every $x \in A$.

The proof is easy once we notice that since g is a fuction $dom\, g = B$ and so range of f is subset of domain of g thus $dom\, g \circ f = A$. Also we observe that $range\, g \circ f \subseteq range\, g$ for any graph. That $F2$ is satisfied is easy.

Definition A.1.19. *A function $f : A \to B$ is invertible if the inverse graph is a function $f^{-1} : B \to A$.*

Theorem A.1.9. A function is invertible iff it is bijective. Furthermore, if a function is invertible then the inverse is a bijective function.

A useful characterization of invertible function is the following:

Theorem A.1.10. A function $f : A \to B$ is invertible iff there is a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$. If such a function g exists then $g = f^{-1}$.

Theorem A.1.11. A function $f : A \to B$ is injective iff there is a function $g : B \to A$ such that $g \circ f = i_A$.

Proof. Let a function $g : B \to A$ be such that $g \circ f(x) = x$ for every $x \in A$. Consider x_1, x_2 such that $f(x_1) = f(x_2)$. Thus $x_1 = (g \circ f)(x_1) = g(f(x_1)) = g(f(x_2)) = x_2$. Hence f is injective. Consider an injective function f and fix an element $a \in A$. Construct $g : B \to A$ as follows. If $y \in range\, f$ let $g(y) = f(x)$. If $y \notin range\, f$ then $g(y) = a$. Easy to see that $g \circ f = i_A$. □

Theorem A.1.12. Let $f : A \to B$ and $g : B \to C$ be functions. Then we can state the following about the composition $g \circ f$,

(1) If f, g are injective then so is $g \circ f$.
(2) If f, g are surjective then so is $g \circ f$.
(3) If f, g are bijective then so is $g \circ f$.

Now we will define Direct and Inverse images of sets under a function.

Definition A.1.20. *Let $f : A \to B$ be a function and consider a set $C \subseteq A$. Then the direct image of C under f is the set of all images of elements of C,*

$$f(C) = \{y \in B \ : \ \exists x \in C \ni f(x) = y\}.$$

Definition A.1.21. *Let $f : A \to B$ be a function and consider a set $D \subseteq B$. Then the inverse image of D under f is the set of all elements in A whose images are elements of D,*

$$f^{-1}(D) = \{x \in A \ : \ \exists y \in D \ni f(x) = y\}.$$

We can alternatively write the inverse image of D as the set

$$f^{-1}(D) = \{x \in A \ : \ f(x) \in D\}.$$

It is important to see how direct images and inverse images act on generalized unions and intersections. The next theorem summarizes there actions.

Theorem A.1.13. Let $f : A \to B$ and let $\{C_i\}_{i\in I}$ and $\{D_i\}_{i\in I}$ be subfamilies in A and B respectively. Then,

(1) $f\left(\bigcup_{i\in I} C_i\right) = \bigcup_{i\in I} f(C_i)$,

(2) $f\left(\bigcap_{i\in I} C_i\right) \subseteq \bigcap_{i\in I} f(C_i)$,

(3) $f^{-1}\left(\bigcup_{i\in I} D_i\right) = \bigcup_{i\in I} f^{-1}(D_i)$,

(4) $f^{-1}\left(\bigcap_{i\in I} D_i\right) = \bigcap_{i\in I} f^{-1}(D_i)$,

Thus, we see that the inverse image is well behaved w.r.t unions and intersections (and complements). Note that the inverse and direct image are functions that maps powersets, i.e $C(\in \mathscr{P}(A)) \mapsto f(C)(\in \mathscr{P}(B))$ and $D(\in \mathscr{P}(B)) \mapsto f^{-1}(D)(\in \mathscr{P}(A))$. This is easy to see if we observe that if $C_1 = C_2$ then $f(C_1) = f(C_2)$. Similarly for inverse images. However, the converse is not true in general.

We have defined the *product* of two classes as $A \times B$ as the class of all the ordered pairs in A and B. We can extend this idea to the product of finite classes $A_1, A_2, \ldots, A_n$ as the class of all *n-tuple* $(a_1, a_2, \ldots, a_n)$ such that $a_i \in A_i$ for all $1 \leq i \leq n$. However, we have a potential problem if we an arbitrary indexed family, $\{A_i \ : \ i \in I\}$. In such a case we have to redefine what a product of classes mean.

Definition A.1.22. *Let $\{A_i : i \in I\}$ be an indexed family of classes; let*
$$A = \bigcup_{i \in I} A_i.$$
The product of the classes A_i is defined to be the class
$$\prod_{i \in I} A_i = \{f : f : I \to A \text{ and } f(i) \in A_i \,\forall\, (i \in I)\}.$$

We will adopt the following notational convention: we designate elements of a product $\prod_{i \in I} A_i$ with boldface letters $\boldsymbol{a}, \boldsymbol{b}$ etc. If $\boldsymbol{a}$ is an element of $\prod_{i \in I} A_i$, we will denote by a_j as $\boldsymbol{a}(j)$. We call a_j as the j^{th} co-ordinate of $\boldsymbol{a}$. Let $A = \prod_{i \in I} A_i$, corresponding to each index we define a function $\Pi_i : A \to A_i$ by $\Pi_i(\boldsymbol{a}) = a_i$. We call Π_i as the $i^{th} - projection$ of A to A_i.

Definition A.1.23. *If A, B are classes, we denote by B^A as the class of all functions whose domain is A and whose co-domain is B.*

In particular if $2 = \{0, 1\}$ denotes the class of two elements, then 2^A is the class of all functions from A to $\{0, 1\}$.

Theorem A.1.14. If A is a set, then $\mathscr{P}(A)$ and 2^A are in $1 - 1$ correspondence.

Proof. We will show that there is a function $f : \mathscr{P}(A) \to 2^A$, such that f is injective. For any $B \in \mathscr{P}(A)$ define $f(B) = \boldsymbol{\chi}_B$. Easy to see that f is injective. Infact there is a bijection. Let $g \in 2^A$. Define $B = g^{-1}(\{1\})$. Then $g = \boldsymbol{\chi}_B$. □

Let us list a couple more axioms that will *almost* complete the set construction axiom.

Axiom 9: If A is a non-empty set, there is an element $a \in A$ such that $a \cap A = \emptyset$.

The above axiom states that a set is disjoint from its elements. Hence if A is a set, the singleton $\{A\} \neq A$.

Axiom 10: If A is a set and $f : A \to B$ is a surjective function, then B is a set.

Next we define relations on sets.

Definition A.1.24. *Let A be a class, by a relation R in A we mean an arbitrary subclass of $A \times A$.*

Let R be a relation in A, then

(1) (Reflexive) R is reflexive if for every $a \in A$, $(a,a) \in R$.

(2) (Irreflexive) R is irreflexive if for every $a \in A$, $(a,a) \notin R$.

(3) (Symmetric) R is symmetric if $(a,b) \in R \implies (b,a) \in R$.

(4) (Asymmetric) R is asymmetric if $(a,b) \in R \implies (b,a) \notin R$.

(5) (Anti-symmetric) R is anti-symmetric if $(a,b),(b,a) \in R \implies a = b$.

(6) (Transitive) R is transitive if $(a,b),(b,c) \in R \implies (a,c) \in R$.

Definition A.1.25. *A relation R in A is called an equivalence relation if it is Reflexive, Transitive and Symmetric.*

Definition A.1.26. *A relation R in A is called a partial order relation if it is Reflexive, Transitive and Anti-symmetric.*

Definition A.1.27. *A relation R in A is called a strict order relation if it is Irreflexive, Transitive and Asymmetric.*

A.2. Countability

In this section we will study in some detail finite sets,countable sets and uncountable sets. We start with finite sets.

Definition A.2.1 (section)**.** *Let n be a positive integer. We use S_n to denote the set of positive integers less than n and call it a section of positive integers. Thus $S_n = \{1, 2, \ldots, n\}$.*

Definition A.2.2 (finite sets)**.** *A set A is said to be finite if there is a bijective correspondence of A with some section of the positive integers, i.e. there exists an $n \in \mathbb{Z}^+$ such that,*

$$f \; : A \to S_n,$$

is a bijective function. We say that A has cardinality n. If A is empty we say that A has cardinality 0.

Remark A.2.1. *S_n itself has cardinality n. We will show that the cardinality of a finite set is uniquely determined by the set.*

Lemma A.2.1. *Let n be a positive integer. Let A be a set and let a_0 be any element of A. Then there exists a bijective correspondence f of the set A with S_{n+1} if and only if there exists a bijective function g from the set $A - \{a_0\}$ with S_n.*

Proof. First assume that $g \; : A - \{a_0\} \to S_n$ is bijective. Let $f \; : A \to S_{n+1}$ be such that $f\big|_{A-\{a_0\}} = g$ and $f(a_0) = n+1$. Then f is bijective.

Assume $f\;: A \to S_{n+1}$ is bijective. We consider two cases.

CASE I: $f(a_0) = n+1$. In this case, define $g = f\big|_{A-\{a_0\}}$. Then we get the desired bijective map.

CASE II: $f(a_0) = k$ where $k \neq n+1$. Then there is an a_1 in A such that $f(a_1) = n+1$. Define $\sigma\;: A \to S_{n+1}$ such that $\sigma(a_0) = n+1$ and $\sigma(a_1) = k$ and for all other values of A, $\sigma = f$. Thus σ is a bijective mapping from A onto S_{n+1} such that $\sigma(a_0) = n+1$. Define $g = \sigma\big|_{A-\{a_0\}}$. Then we get the desired bijective map. $\square$

Theorem A.2.1. Let A be a set; suppose there exists a bijection $f\;: A \to S_n$ for some $n \in \mathbb{Z}^+$. Let B be a proper subset of A. Then there exists no bijection $g\;: A \to S_n$, but if B is not empty, then there is a bijection $f\;: B \to S_m$ for some $m < n$.

Proof. We will prove this by induction on n. First note that when A is empty, there is no proper subset of A and so the theorem is trivially true. Let $f\;: A \to S_1$ be a bijection. Then any proper subset B of A is empty and so cardinality of $B = 0$. Assume the theorem is true for some n in $\mathbb{Z}^+$. Let $f\;: A \to S_{n+1}$ be a bijection and let B be a proper subset of A. Assume B is not empty. There is an element a_0 in B and note that a_0 is also in A. By the previous Lemma, there is a map,

$$g\;: A - \{a_0\} \to S_n,$$

such that g is bijective. Note that $B - \{a_0\}$ is a proper subset of $A - \{a_0\}$. Now we can use our inductive hypothesis to state that:

(1) There is no bijection $h\;: B - \{a_0\} \to S_n$.
(2) Either $B - \{a_0\}$ is empty or there is an $m < n$ such that, $k\;: B - \{a_0\} \to S_m$ is bijective.

Using the Lemma above we can state (1) equivalently by stating that there is not bijection between B and S_{n+1}. If $B - \{a_0\}$ is not empty, then we can use the previous lemma to state that there is a bijection from B to S_{m+1} where $m+1 < n+1$ and so we have proved the statement for $n+1$. $\square$

Corollary A.2.1.1. *If A is finite, there is no bijection of A with a proper subset of itself.*

Proof. Let $f\;: B \to A$ be bijective. Since A is finite there is a bijective function $g\;: A \to S_n$. Define $h = g \circ f$. Then $h\;: B \to S_n$ is bijective which contradicts the theorem above. $\square$

Corollary A.2.1.2. $\mathbb{Z}^+$ *is not finite.*

Proof. We prove this by the contrapositive of the above corollary i.e. we construct a bijection between $\mathbb{Z}^+$ and a propoer subset $\mathbb{Z}^+ - \{1\}$. Let f : $\mathbb{Z}^+ - \{1\} \to \mathbb{Z}^+$ be given by $f(i) = i - 1$. Then f is bijective. □

Corollary A.2.1.3. *The cardinality of a finite set A is uniquely determined by A i.e. there doesnot exist S_n, S_m with $n \neq m$ such that A is in bijective correspondence with both S_n, S_m.*

Proof. Suppose there exist bijections f, g from A onto S_n, S_m where $n \neq m$. Without loss of generality, assume $n < m$. Let $h = g \circ f^{-1}$. Then $h : S_n \to S_m$ is bijective which gives a contradiction since S_n is a proper subset of S_m. □

Corollary A.2.1.4. *If B is a subset of the finite set A, then B is finite. If B is a proper subset of A, then the cardinality of B is less than A.*

Corollary A.2.1.5. *Let B be a non-empty set. Then the following are equivalent:*

(1) B is finite.

(2) There is a surjective function from a section of the positive integers onto B.

(3) There is an injective function from B into a section of the positive integers.

Proof. We will prove (1) $\implies$ (2), (2) $\implies$ (3) and (3) $\implies$ (1).

(1) $\implies$ (2). Assume B is finite. Then, for some $n \in \mathbb{Z}^+$, there is a bijective map $f : B \to S_n$. Thus $f^{-1} : S_n \to B$ is onto.

(2) $\implies$ (3). Assume there is a map $f : S_n \to B$ such that f is onto. Define the function $h : B \to S_n$ where for any $b \in B$, $h(b) = \min f^{-1}(b)$. Since f is onto $f^{-1}(b)$ is not empty and is a subset of S_n and by the well ordering property of integers, the minimum exists and is unique. Thus h is injective.

(3) $\implies$ (1). Assume there is an injective map $f : B \to S_n$ for some $n \in \mathbb{Z}^+$. Then $f : B \to f(B)$ is bijective. Also, since $f(B) \subset S_n$, it is finite. Hence B is finite. □

Theorem A.2.2. Finite Unions and finite cartesian products of finite sets are finite.

Proof. Let $\mathscr{B} = \{B_i : 1 \le i \le n\}$ be a finite collection of finite sets. We will show that $\bigcup_{i=1}^{n} B_i$ is also finite by induction on n. When $n = 1$, the result is trivial. Consider the case when $n = 2$. By hpyothesis, there exists bijective functions f_1, f_2 from B_1, B_2 onto S_{n_1}, S_{n_2} for some n_1, n_2 in $\mathbb{Z}^+$.

Define $h : S_{n_1+n_2} \to B_1 \cup B_2$ as follows: $h(i) = f_1(i)$ for $1 \le i \le n_1$ and $h(n_1 + j) = f_2(j)$ for $1 \le j \le n_2$. Thus h is a surjection from a section of integers to $B_1 \cup B_2$ and hence, $B_1 \cup B_2$ is finite. Note that h is note necessarily a bijection since A, B may not be disjoint. The result now follows from induction.

Note that for any $b \in B_1$, $\{b\} \times B_2$ is finite because the function $f : \{b\} \times B_2 \to S_{n_2}$ given by $f((b,c)) = f_2(c)$ for any $c \in B_2$ is a bijection. Now, $B_1 \times B_2$ can be written as $\bigcup_{b \in B_1} \{b\} \times B_1$, which is a finite union of finite sets and so is finite. The result for arbitrary n follows from induction. □

Just as sections of integers are the prototypes for finite sets, the set of all positive integers is prototype for the simplest infinite set that we call countably infinite or denumerable sets.

Definition A.2.3 (Infinite and Countably infinite)**.** *A set A is said to be infinite if it is not finite. It is said to be countably infinite if there is a bijective correspondence,*

$$f : A \to \mathbb{Z}^+.$$

From all our observations regarding finite sets we can immediately deduce the following about infinite sets

(1) For any proper subset B of A, if B is infinite then A is also infinite.

(2) If there is a proper subset B of A and if there is a function $f : B \to A$ which is bijective, then A is infinite.

Definition A.2.4 (Countable)**.** *A set A is said to be countable if it is either finite or if it is countably infinite. A set that is not countable is said to be uncountable.*

We saw that $\mathbb{Z}^+$ was not finite. It is countably infinite by definition and hence $\mathbb{Z}^+$ is countable. Is $\mathbb{Z}$ countable?

Proposition A.2.1. *$\mathbb{Z}$ is countably infinite.*

Proof. $\mathbb{Z}^+$ is a proper subset of $\mathbb{Z}$ and is not finite. Hence we must show that there is a function $f : \mathbb{Z} \to \mathbb{Z}^+$ that is bijective. Let us define,

$$f(n) = \begin{cases} -2n & \text{if } n < 0, \\ 2n+1 & \text{if } n \ge 0. \end{cases}$$

Easy to see that f is a bijection. □

Proposition A.2.2. *The product $\mathbb{Z}^+ \times \mathbb{Z}^+$ is countably infinite.*

Proof. The proof is the famous diagonal arrangement. If we of $\mathbb{Z}^+ \times \mathbb{Z}^+$ as a grid, we can start counting **diagonally** from right to top. Let $f : \mathbb{Z}^+ \to \mathbb{Z}^+$

be given by, $f(i,j) = j + \sum_{k=1}^{i+j-2} k$. Thus, for example, $f(1,1) = 1$, $f(2,1) = 2$, $f(1,2) = 3$ and so on. Easy to see that we have a bijection. For example to show it is surjective let m be any positive integer, say $m = 11$. Then think of 11 as a list $(1), (2,3), (4,5,6), (7,8,9,10), (11,12,13,14,15)$. Thus 11 is on the 5^{th} list which means that j will be such that $1 \leq j \leq 5$. Similarly i will be less than 5. Easy to see that $i = 5$ and $j = 1$ will give $f(i,j) = 11$. $\square$

There is a very useful criterion for showing that a set is countable.

Theorem A.2.3. Let B be a non-empty set. Then the following are equivalent.

(1) B is countable.
(2) There is a surjective function $f : \mathbb{Z}^+ \to B$.
(3) There is an injective function $g : B \to \mathbb{Z}^+$.

Proof. Assume that B is countable. Hence B is either finite or countably infinite. In the later case, there is a surjection $f : \mathbb{Z}^+ \to B$. If B is finite, then there is a bijection $g : S_n \to B$ for some $n \in \mathbb{Z}^+$. Define $f : \mathbb{Z}^+ \to B$ as follows: $f(i) = g(i)$ for $i \leq n$ and $f(i) = g(n)$ for $i > n$. Thus f is a surjection. Hence we have shown $(1) \implies (2)$.

Assume that there is a surjection $f : \mathbb{Z}^+ \to B$. This means that for any b in B, $f^{-1}(\{b\}) \neq \emptyset$. Moreover, for $b_1, b_2 \in B$, if $b_1 \neq b_2$, then $f^{-1}(\{b_1\}) \neq f^{-1}(\{b_2\})$. Let $g(b) = \min f^{-1}(\{B\})$. Then $g(b)$ is an injective function from B to $\mathbb{Z}^+$. Thus, we have shown that $(2) \implies (3)$.

Assume there is an injective function $g : B \to \mathbb{Z}^+$. To show that B is countable we need to show that either B is finite or B is countably infinite. We can assume B is not finite. Thus $g(B) \subset \mathbb{Z}^+$ is not finite. IF $g(B)$ was countably infinite, then B would be countably infinite. Hence we need to show that any infinite subset of $\mathbb{Z}^+$ is countably infinite which would mean that $(3) \implies (1)$. $\square$

Remark A.2.2. *The above theorem (part (2)) states that for a countable set we can list its elements as a* ***sequence****. Thus if A is countable, then $A = \{a_1, a_2, \cdots\}$.*

Proposition A.2.3. *Any infinite subset of $\mathbb{Z}^+$ is countably infinite.*

Proof. Let B be an infinite subset of $\mathbb{Z}^+$. There is bijection $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ that lists the elements of $\mathbb{Z}^+$ in a sequence. The intuitive idea is to go along the sequence and check if it belongs to B. We want a function $g : \mathbb{Z}^+ \to B$ that is bijective. For example as we move along the sequence in f, the first occurance of an element in B occured in the index k. We set $g(1) = f(k)$. Then we continue traversing along the sequence in f until we find an index

$l > k$ such that $f(l) \in B$. We set $g(2) = f(l)$. It is clear that such a function is surjective. How do we define the function g? Let us tweak g a bit. Instead of taking the first occurance of an element listed by f, we take the smallest element of B listed by f. There must be an index m such that $f(m) = \min B$. We take $g(1) = f(m)$ and proceed subsequently.

Let $g(1)$ be the smallest element that is in B. This is possible since $B \subset \mathbb{Z}^+$. Let $g(2)$ be the smallest element other than $g(1)$ that is in B. This is called an **inductive** definition (or recursion). We define a function whose domain is the natural numbers inductively, in this case,

$$g(n) = \min \{B - \{g(1), g(2), \ldots, g(n-1)\}\}.$$

It is clear that g is injective. To see that g is surjective, consider an arbitrary element, say b, of B. For any $m \in \mathbb{Z}^+$ there is an $n \in \mathbb{Z}^+$ such that $g(n) > S_m$. Because if not, then B would be finite. In particular for $m = b$ there is an n such that $g(n) > b$. Let $C = \{n \in \mathbb{Z}^+ \ : \ g(n) \geq b\}$. Then $C \subset \mathbb{Z}^+$ and so a minimum $p \in \mathbb{Z}^+$ exists such that $g(p) \geq b$. This means that $g(1), g(2), \ldots, g(p-1)$ are all less than b. By definition of g, $g(p) \leq b$. Thus it must be the case that $g(p) = b$. $\square$

We have used a key idea here called the principle of recursive definition.

Definition A.2.5 (Principle of recursive definition)**.** *Let A be any set. Given a formula that defines $g(1)$ as a **unique** element of A and for all $i > 1$ defines $g(1)$ **uniquely** as an element of A in terms of the values of g for positive integers less than i, this formula determines a **unique function** $g : \mathbb{Z}^+ \to A$.*

Corollary A.2.3.1. *A subset of a countable set is countable.*

Proof. Let $A \subset B$ such that B is countable. Thus there exists an injection $f : B \to \mathbb{Z}^+$. Hence, $f\big|_A : A \to \mathbb{Z}^+$ is also an injection. $\square$

Corollary A.2.3.2. *The set $\mathbb{Z}^+ \times \mathbb{Z}^+$ is countable.*

Proof. Let $f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+$ be given by $f(i,j) = 2^i 3^j$. Then, f is injective (because $2, 3$ are prime). $\square$

Corollary A.2.3.3. *The set of positive rational numbers is countable.*

Proof. Let $f : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Q}^+$ be defined by $f(i,j) = \frac{i}{j}$. Then f is surjective. We showed that there is a function $g : \mathbb{Z}^+ \to \mathbb{Z}^+ \times \mathbb{Z}^+$ that is surjective. Thus $f \circ g : \mathbb{Z}^+ \to \mathbb{Q}^+$ is surjective. $\square$

Theorem A.2.4. A countable union of countable sets is countable.

Proof. Let $\mathscr{A}$ be a collection of countable sets $A_1, A_2, \ldots$. Since each A_i is countable there is a function $f_i : \mathbb{Z}^+ \to A_i$ that is a surjection. In other words we can list the elements of A_i's,

$$\begin{aligned} A_1 &= \{a_{11}, a_{12}, a_{13}, \ldots\} \\ A_2 &= \{a_{21}, a_{22}, a_{23}, \ldots\} \\ &\vdots \\ A_j &= \{a_{j1}, a_{j2}, a_{j3}, \ldots\} \\ &\vdots \end{aligned}$$

Define $g : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \bigcup_{i=1}^{\infty} A_i$ as, $g(i,j) = a_{ij} = f_i(j)$. Thus g is a surjection. $\square$

Theorem A.2.5. A finite product of countable sets is countable.

Proof. We have shown the analogous result for finite sets. Let $\{A_i : 1 \le i \le n\}$ be a finite collection of countable sets. We prove this by induction on n. The case $n = 1$ is trivial. When $n = 2$, $A_1 \times A_2 = \bigcup_{a \in A_1} \{a\} \times A_2$ is a countable unioon of countable sets and is countable by the theorem above. We proceed by induction to prove it for any $n \in \mathbb{Z}^+$. $\square$

Theorem A.2.6. Let A be any set. There is NO injective map $f : \mathscr{P}(A) \to A$ and there is no surjective map $g : A \to \mathscr{P}(A)$.

Proof. We will show that there is no surjective map which maps A onto the powerset of A. This will then imply that there is no injective map that maps powerset of A into A. This is because of the following:

$$\exists\,(f : B \to C)\, f \text{ is injective, then } \exists\,(g : C \to B)\, g \text{ is surjective.}$$

Taking the contrapositive of this statement we can just show the non-existence of a surjective map. To see why the above statement is true, assume there is an injection $f : B \to C$. Thus for any $c \in f(B)$, there is a unique b such that $f(b) = c$. Define $g : C \to B$ as follows, if $c \in C$ is in $f(B)$, then $g(c) = b$ such that $f(b) = c$. If $c \in C$ such that $c \notin f(B)$, then fix a $b_0 \in B$ and let $g(c) = b_0$. Thus g is surjective. Note that B has to be non-empty for all this to make sense.

Now we will show the non-existence of a surjective map from A onto $\mathscr{P}(A)$ by contradiction. Assume, there is a $g : A \to \mathscr{P}(A)$ such that g is surjective. Thus for any $X \subset A$, there is a $a \in A$ such that $g(a) = X$. Let $B = \{a \in A : a \in g(a)\}$. Then $B \subset A$ and hence by our assumption there is an $a_0 \in A$ such that $g(a_0) = B$. There are two possibilities for a_0. Either

$a_0 \in g(a_0)$ or $a_0 \notin g(a_0)$. In the first case, $a_0 \notin B$ and in the second case $a_0 \in B$. Thus $g(a_0) \neq B$. Hence, g cannot be surjective. □

Corollary A.2.6.1. *There exists an uncountable set.*

Proof. There is no surjection from $\mathbb{Z}^+$ onto $\mathscr{P}(\mathbb{Z}^+)$. Thus, $\mathscr{P}(\mathbb{Z}^+)$ is uncountable. □

Let us now give a characterization of infinite sets. These sets are either countably infinite or uncountable. From our discussion following the definition of infinite sets, we already gathered a few notions of what is means to be for a set to be infinite. We elaborate on this now.

Theorem A.2.7. Let A be a set. The following statements about A are equivalent.

(1) There is an injective function $f : \mathbb{Z}^+ \to A$.
(2) There exists a bijection of A with a proper subset of itself.
(3) A is infinite.

Proof. (1) $\implies$ (2) Let $f : \mathbb{Z}^+ \to A$ be an injective map. Note that $f : \mathbb{Z}^+ \to f(\mathbb{Z}^+)$ is a bijective map that lists a subset of A as a sequence $(a_1, a_2, \cdots)$. Let $B = A - \{a_1\}$. Then B is a proper subset of A. Note that $A = f(\mathbb{Z}^+) \cup (A - f(\mathbb{Z}^+))$. Define $g : A \to B$ as follows:

$$g(a) = \begin{cases} a_{n+1} & \text{if } a = x_n \in f(\mathbb{Z}^+) \text{ for some } n \\ a & \text{if } a \notin f(\mathbb{Z}^+) \end{cases}$$

(2) $\implies$ (3) This just follows from our discussion following the definition of infinite sets.

(3) $\implies$ (1) First pick **any** element of A and call it a_1. Thus $f(1) = a_1$. Now, pick any element of A which is not equal to a_1 and call it a_2. That is $f(2) \in A - \{a_1\}$. This process will continue indefinitely since A is infinite. Hence $f : \mathbb{Z}^+ \to A$ is injective, where f is definied recursively as follows:

$$f(n) \in A - \{a_1, a_2, \ldots, a_n\}.$$

However, there is a problem in our recursive definition for f. Namely, $f(1)$ is not unique and $f(n)$ is certainly not uniquely defined in terms of $\{a_1, a_2, \ldots, a_n\}$, since we are picking an arbitrary element of A. We need an additional axiom from set theory to make this possible. □

Axiom 11: (Axiom of choice) Given a collection $\mathscr{B}$ of non-empty sets, there exists a function

$$c : \mathscr{B} \to \bigcup_{B \in \mathscr{B}} B,$$

such that $c(B) \in B$ for any $B \in \mathscr{B}$.

Using the axiom of choice we can fix our recursive definition for f. Let $\mathscr{B} = \mathscr{P}(A) - \emptyset$. Then there exists a choice function such that $c(B) \in B$ for any $B \in \mathscr{B}$. Pick $B = A$. Thus $c(A)$ gives an element of A uniquely. Let $f(1) = c(A)$. Now $A - \{f(1)\}$ is also in $\mathscr{B}$. Define $f(2) = c(A - \{f(1)\})$ which is again uniquely defined. Thus, define

$$f(n) = c(A - \{f(1), f(2), \ldots, f(n-1)\}).$$

This is a valid recursive function.

There is a equivalence relation which is induced by bijective functions. Let us define $A \sim B$ whenever there is a bijective map $f : A \to B$. It is easy to see that $\sim$ is an equivalence relation. A set A is countably infinite if $A \sim \mathbb{Z}^+$. Note that any two countably infinite sets are equivalent since if $A \sim \mathbb{Z}^+$ and $B \sim \mathbb{Z}^+$, then $A \sim B$.

This is an extremely important idea which we will explore now in the subsequent paragraph in the form of Cantor-Schroder-Bernstein theorem.

Suppose A, B are sets and f is one-to-one function from A into B. Then $A \sim f(A) \subset B$, so it is natural to think of B as being at least as large as A. This suggests the following notation:

Definition A.2.6. *If A, B are sets, then we will say that B* ***dominates*** *A, and write $A \preceq B$, if there is an injective function $f : A \to B$. If $A \preceq B$ and $A \not\sim B$, then we say B strictly dominates A.*

Theorem A.2.8 (Schroder-Bernstein theorem)**.** If $A \preceq B$ and $B \preceq A$, then $A \sim B$.

Proof. Let $f : A \to B$ and $g : B \to C$ be injective functions. Note that $g : B \to g(B)$ is bijective. Let $A_0 = A - g(B)$ and let us define recursively $A_n = g(f(A_{n-1}))$. Let $X = \bigcup_{n=0}^{\infty} A_n$. Then $X \subset A$. Define $Y = A - X$. Note that, $Y \subset g(B)$. This is because if $a \in Y$ then $a \in A$ and $a \notin A_0$, that is $a \in g(B)$. Define,

$$h(a) = \begin{cases} f(a) & \text{if } a \in X \\ g^{-1}(a) & \text{if } a \in Y \end{cases}$$

Claim: h is injective. Assume $h(a_1) = h(a_2)$ for a_1, a_2 in A. The only cases we have to check are when $a_1 \in X$ and $a_2 \in Y$ and vice-versa. Suppose $a_1 \in X$ and $a_2 \in Y$. Then $f(a_1) = g^{-1}(a_2)$. Which means that $g(f(a_1)) = a_2$. Since $a_1 \in X$, there is an n such that $a_1 \in A_n$ and hence $a_2 \in g(f(A_n))$ which means that $a_2 \in A_{n+1}$ i.e. $a_2 \in X$ which is a contradiction. The other case is similar.

Claim: h is surjective. Let $b \in B$. Then $g(b) \in A$. Thus $g(b) \in X$ or $g(b) \in Y$. If $g(b)$ is in Y, then $h(g(b)) = g^{-1}(g(b)) = b$. Thus there is an $a = g(b) \in A$ such that $h(a) = b$. If $g(b) \in X$, then there is an A_n such that $g(b) \in A_n$. But n cannot be 0. Thus there is an $n > 1$ such that $g(b) \in g(f(A_{n-1}))$. This means that there is an $x_1 \in f(A_{n-1})$ such that $g(x_1) = g(b)$. Since g is injective, this means that $x_1 = b$ i.e. there is an $x \in A_{n-1}$ such that $f(x) = x_1 = b$. But this means that $h(x) = b$ for some $x \in A_{n-1} \subset X$. $\square$