Measure Theory for Applications to Probability

Seung Hyun Kim

August 17, 2024

Chapter 1

Topological Preliminaries

1.1 Topologies and Bases

Let X be an arbitrary set. Then, a topology τ on X is a collection of subsets of X satisfying the following conditions:

i) $\emptyset, X \in \tau$; τ contains the empty set and the entire set

ii) For any $A, B \in \tau$, $A \cap B \in \tau$ as well; τ is closed under finite intersections

iii) For any collection $\{A_\alpha\} \in \tau$, $\bigcup_\alpha A_\alpha \in \tau$; τ is closed under arbitrary unions

The elements of τ are called open sets, and the pair (X, τ) is referred to as a topological space. Any set whose complement is open is called a closed set.

A base of X is a collection $\mathbb{B}$ of subsets of X such that:

i) $\mathbb{B}$ covers X, that is, $X \subset \bigcup_{B \in \mathbb{B}} B$

ii) For any $B_1, B_2 \in \mathbb{B}$ and $x \in B_1 \cap B_2$, there exists a $B_3 \in \mathbb{B}$ such that $x \in B_3 \subset B_1 \cap B_2$.

Let $\mathbb{B}$ be a base of X. Consider the collection τ of subsets of X satisfying the following properties:

i) $\emptyset \in \tau$

ii) τ contains any set $U \subset X$ such that, for any $x \in U$, there exists a $B \in \mathbb{B}$ such that $x \in B \subset U$

Then, τ is a topology on X, and it is called the topology generated by the base $\mathbb{B}$. Clearly, $\mathbb{B}$ is contained in τ. We first show that the collection defined above is a topology:

Lemma 1.1 τ is a topology.

Proof) We show that τ satisfies all three conditions of a topology.

i) $\emptyset \in \tau$ by definition. Since $X \subset \bigcup_{B \in \mathbb{B}} B$, for any $x \in X$ there exists a $B \in \mathbb{B}$ such that $x \in B$, and because B is a subset of X, $x \in B \subset X$. Therefore, $X \in \tau$ as well.

ii) For any $A, B \in \tau$, if $A \cap B = \emptyset$, $A \cap B \in \tau$.
If $A \cap B \neq \emptyset$, for any $x \in A \cap B$, $x \in A$ and $x \in B$, so there exist $A_1, B_1 \in \mathbb{B}$ such that $x \in A_1 \subset A$ and $x \in B_1 \subset B$.
By the property of a base of X, since $x \in A_1 \cap B_1$ and $A_1, B_1 \in \mathbb{B}$, there exists a $C \in \mathbb{B}$ such that $x \in C \subset A_1 \cap B_1$.
Finally, since $A_1 \cap B_1 \subset A \cap B$, we have $x \in C \subset A \cap B$. This proves that $A \cap B \in \tau$.

iii) For any collection $\{A_\alpha\} \subset \tau$, denote $A = \bigcup_\alpha A_\alpha$. If $A = \emptyset$, then $A \in \tau$ trivially.
Suppose $A \neq \emptyset$, and choose $x \in A$. Then, $x \in A_a$ for some a, and because $A_a \in \tau$, there exists a $B \in \mathbb{B}$ such that $x \in B \subset A_a$. $A_a \subset A$, so $x \in B \subset A$; this shows that $A \in \tau$.

Therefore, τ is a topology on X.
Q.E.D.

A converse argument can also be made.

Lemma 1.2 Let τ be a topology on X, and $\mathbb{B}$ a collection of subsets of E. Suppose that:

i) $\mathbb{B}$ is contained in τ, that is, any subset in $\mathbb{B}$ is also in τ

ii) For any nonempty $A \in \tau$ and $x \in A$, there exists a $B \in \mathbb{B}$ such that $x \in B \subset A$.

Then, $\mathbb{B}$ is a base on X that generates τ.

Proof) The proof proceeds in two steps. First, we must prove that $\mathbb{B}$ is a base of X. Then, we must prove that $\mathbb{B}$ generates τ.

$X \in \tau$ by the definition of a topology. For any $x \in X$, by hypothesis there exists a $B_1 \in \mathbb{B}$ such that $x \in B_1 \subset \bigcup_{B \in \mathbb{B}} B$; this holds for any $x \in X$, so $X \subset \bigcup_{B \in \mathbb{B}} B$.
Now let $B_1, B_2 \in \mathbb{B}$ and consider $x \in B_1 \cap B_2$. Since $\mathbb{B}$ is contained in τ and topologies are closed under finite intersections, $B_1 \cap B_2 \in \tau$. By hypothesis, this means that there exists a $B \in \mathbb{B}$ such that $x \in B \subset B_1 \cap B_2$.
Therefore, $\mathbb{B}$ is a base of X.

It remains to show that $\mathbb{B}$ generates τ. However, this fact follows directly from the hypothesis and the fact that any topology contains the empty set.

Q.E.D.

Note that to each base there only exists a single topology that it generates (defined above), while there can be mutliple bases that generate a single topology.

There are two other ways to characterize the topology generated by a base that will prove to be much more useful. In fact, one characterization shows us how to construct any open set in a topology given a base that generates it. The other characterization is purely abstract, and is stated to emphasize the similiarities between topologies and σ-algebras.

Theorem 1.3 Let τ be a topology on X and $\mathbb{B}$ a base of X. Then, the following are equivalent:

i) τ is the topology generated by $\mathbb{B}$.

ii) $\mathbb{B}$ is contained in τ and, for any $B \in \tau$, there exist a collection $\{B_i\} \subset \mathbb{B}$ such that $B = \bigcup_i B_i$; any set in τ can be represented as an arbitrary union of sets in $\mathbb{B}$.

iii) τ is the smallest topology containing $\mathbb{B}$, that is, τ is the intersection of every toplogy on X that contains $\mathbb{B}$.

Proof) We first show that *i)* implies *ii)*. Let τ be the topology generated by $\mathbb{B}$. Then, $\mathbb{B}$ is trivially contained in τ.
For any $B \in \tau$, if $B = \emptyset$, then $B \subset B_1$ for any $B_1 \in \mathbb{B}$. If $B \neq \emptyset$, then for any $x \in B$, there exists a $B_x \in \mathbb{B}$ such that $x \in B_x \subset B$ by the definition of the topology generated by $\mathbb{B}$.
This holds for any $x \in B$, so we can see that

$$B \subset \bigcup_{x \in B} B_x.$$

On the other hand, since each B_x is a subset of B,

$$\bigcup_{x \in B} B_x \subset B.$$

It follows that $B = \bigcup_{x \in B} B_x$, so that B can be expressed as the union of sets in $\mathbb{B}$.

Now we show that *ii)* implies *iii)*. Suppose that $\mathbb{B} \subset \tau$ and that any $B \in \tau$ can be expressed as the union of sets in $\mathbb{B}$. Let $\{\tau_s\}$ be the collection of all topologies on X that contain $\mathbb{B}$, and choose any topology τ_s that contains $\mathbb{B}$.
For any $A \in \tau$, because $A = \bigcup_i B_i$ for a collection $\{B_i\} \subset \mathbb{B}$, and each $B_i \in \tau_s$ because $\mathbb{B} \subset \tau_s$, it follows from the closedness of topologies under arbitrary unions that

$$A = \bigcup_i B_i \in \tau_s.$$

This holds for any $A \in \tau$, so $\tau \subset \tau_s$. This in turn holds for any topology τ_s that contains $\mathbb{B}$, so

$$\tau \subset \bigcap_s \tau_s.$$

Finally, since τ is also a topology on X that contains $\mathbb{B}$, τ is one of the topologies included in the colleciton $\{\tau_s\}$ and thus

$$\bigcap_s \tau_s \subset \tau.$$

It follows that $\tau = \bigcap_s \tau_s$.

Lastly, we can show that *iii*) implies *i*). Let τ be the smallest topology on X containing $\mathbb{B}$, and denote by s the topology generated by $\mathbb{B}$. It will be shown that $\tau = s$.
It is immediately clear that $\tau \subset s$, since s is a topology on X containing $\mathbb{B}$. To see the reverse inclusion, let $A \in s$. Because we have already shown that $i) \to ii)$, and s is the topology generated by $\mathbb{B}$, A is the union of sets in $\mathbb{B}$. But $\mathbb{B}$ is contained in τ and topologies are closed under arbitrary unions, so it must be the case that $A \in \tau$. This holds for any $A \in s$, so $s \subset \tau$ and $\tau = s$, that is, τ is the topology generated by $\mathbb{B}$.
Q.E.D.

It was shown above that the topology generated by a base can be represented as an arbitrary intersection of topologies. It is also the case that any arbitrary intersection of topologies is also a topology, as shown in the next result.

Lemma 1.4 The intersection of topologies is also a topology.

Proof) Let $\{\tau_\alpha\}$ be an arbitrary collection of topologies on X, and define $\tau = \bigcap_\alpha \tau_\alpha$. We show that τ satisfies the three conditions for a topology.

i) $\emptyset, X \in \tau_\alpha$ for all α, so they are contained in τ as well.

ii) For any $B_1, B_2 \in \tau$, because $B_1, B_2 \in \tau_\alpha$ for all α as well, $B_1 \cap B_2 \in \tau_\alpha$ since all the τ_α are topologies on X. This implies that $B_1 \cap B_2 \in \tau$.

iii) Let $\{A_i\} \subset \tau$. Then, $\{A_i\} \subset \tau_\alpha$ for all α as well, so that, by the properties of a topology, $\bigcup_i A_i \in \tau_\alpha$ for all α. By implication, $\bigcup_i A_i \in \tau$.

Therefore, τ is a topology.
Q.E.D.

1.2 Order Topologies

A special kind of topology which we will often encounter are order topologies. Suppose that E is a totally ordered set such as the real line or the extended real line. Then, the order topology on E is that generated by the base $\mathbb{B}$ that consists of rays

$$\{x \in E \mid x < a\}, \quad \{x \in E \mid x > b\},$$

and the open intervals

$$\{x \in E \mid a < x < b\}$$

for any $a, b \in E$, provided that E has at least two elements.
In order for the above definition to make sense, $\mathbb{B}$ must be a base on X; this fact can be easily but tediously shown, so the proof is omitted.

1.2.1 The Standard Topologies on $\mathbb{R}$ and $[-\infty, +\infty]$

The main order toplogies that we will use are those on $\mathbb{R}$ and $\bar{\mathbb{R}} = [-\infty, +\infty]$. For the former, the order topology $\tau_{\mathbb{R}}$ on $\mathbb{R}$ is the topology generated by the base

$$\mathbb{B} = \{(-\infty, a) \mid a \in \mathbb{R}\} \bigcup \{(a, +\infty) \mid a \in \mathbb{R}\} \bigcup \{(a, b) \mid a, b \in \mathbb{R}\},$$

while for the latter, the order topology $\tau_{[-\infty,+\infty]}$ on $[-\infty, +\infty]$ is the one generated by the base

$$\bar{\mathbb{B}} = \{[-\infty, a) \mid a \in \mathbb{R}\} \bigcup \{(a, +\infty] \mid a \in \mathbb{R}\} \bigcup \{(a, b) \mid a, b \in \mathbb{R}\}.$$

Note that $\tau_{\mathbb{R}}$ is also generated by the base (it is easily shown that the collection defined below is actually a base) given as

$$\mathbb{B}' = \{(a, b) \mid a, b \in \mathbb{R}\};$$

for any nonempty $A \in \tau_{\mathbb{R}}$, because $\tau_{\mathbb{R}}$ is generated by $\mathbb{B}$, by theorem 1.3 there exists a collection $\{B_i\} \subset \mathbb{B}$ such that $A = \bigcup_i B_i$. However, any $B \in \mathbb{B}$ can in turn be represented as the union of open intervals, since $(-\infty, a) = \bigcup_{n \in N_+} (a - n, a)$ and $(a, +\infty) = \bigcup_{n \in N_+} (a, a + n)$, so A can also be expressed as the union of open intervals. $\mathbb{B}'$ is contained in $\tau_{\mathbb{R}}$ because $\mathbb{B}$ is, and therefore, by theorem 1.3, $\tau_{\mathbb{R}}$ is generated by the base $\mathbb{B}'$.
We call the topologies $\tau_{\mathbb{R}}$ and $\tau_{[-\infty,+\infty]}$ the standard topologies on $\mathbb{R}$ and $[-\infty, +\infty]$. Their relationships, as well as the relationships between the standard topology and other topologies on $\mathbb{R}$, will be considered later on.

It is worth noting at this point that there actually exist countable bases that generate the standard topology on $\mathbb{R}$ and $[-\infty,+\infty]$. Specifically, letting $\mathbb{Q}$ be the set of all rational numbers, $\tau_{\mathbb{R}}$ and $\tau_{[-\infty,+\infty]}$ are generated by the countable bases

$$\mathcal{B} = \{(a,b) \mid a,b \in \mathbb{Q}\}$$

and

$$\bar{\mathcal{B}} = \{[-\infty,a) \mid a \in \mathbb{Q}\} \bigcup \{(a,+\infty] \mid a \in \mathbb{Q}\} \bigcup \{(a,b) \mid a,b \in \mathbb{Q}\}.$$

To show this, we can make use of lemma 1.2 and theorem 1.3 as follows:

Theorem 1.5 The standard topologies on $\mathbb{R}$ and $[-\infty,+\infty]$ are generated by the countable bases $\mathcal{B}$ and $\bar{\mathcal{B}}$.

Proof) We first consider $\tau_{\mathbb{R}}$. Since $\mathcal{B} \subset \mathbb{B}' \subset \tau_{\mathbb{R}}$, $\tau_{\mathbb{R}}$ contains the base $\mathcal{B}$.
Now let $A \in \tau_{\mathbb{R}}$ be a nonempty set and choose any $x \in A$. Because $\mathbb{B}'$ is a base generating $\tau_{\mathbb{R}}$, there exists an $(a,b) \in \mathbb{B}'$ such that

$$x \in (a,b) \subset A.$$

Because $\mathbb{Q}$ is dense in $\mathbb{R}$ and $a < x < b$, there exist $q,r \in \mathbb{Q}$ such that $a < q < x < r < b$. This indicates that

$$x \in (q,r) \subset (a,b) \subset A,$$

where $(q,r) \in \mathcal{B}$. Therefore, by lemma 1.2, $\mathcal{B}$ generates $\tau_{\mathbb{R}}$.

To show that $\tau_{[-\infty,+\infty]}$ is generated by $\bar{\mathcal{B}}$, we proceed in much the same way. $\bar{\mathcal{B}} \subset \bar{\mathbb{B}} \subset \tau_{[-\infty,+\infty]}$ tells us that $\tau_{[-\infty,+\infty]}$ contains the base $\bar{\mathcal{B}}$.
For any nonempty $A \in \tau_{[-\infty,+\infty]}$ and $x \in A$, because $\bar{\mathbb{B}}$ is a base generating $\tau_{[-\infty,+\infty]}$, there exists a $B \in \bar{\mathbb{B}}$ such that

$$x \in B \subset A.$$

We now consider three separate cases:

i) $B = (a,b)$ for some $a,b \in \mathbb{R}$
In this case, the denseness of $\mathbb{Q}$ in $\mathbb{R}$ shows, as above, that there exists a $(q,r) \in \bar{\mathcal{B}}$ such that $x \in (q,r) \subset B \subset A$.

ii) $B = [-\infty,a)$ for some $a \in \mathbb{R}$

In this case, choosing $q \in \mathbb{Q}$ such that $q < a$, it follows that $[-\infty, q) \in \bar{\mathcal{B}}$ and

$$x \in [-\infty, q) \subset B \subset A.$$

iii) $B = (a, +\infty]$ for some $a \in \mathbb{R}$
In this case, choosing $q \in \mathbb{Q}$ such that $a < q$, it follows that $(q, +\infty] \in \bar{\mathcal{B}}$ and

$$x \in (q, +\infty] \subset B \subset A.$$

Thus, in any case, there exists a $B' \in \bar{\mathcal{B}}$ such that

$$x \in B' \subset B \subset A.$$

Therefore, by lemma 1.2, $\bar{\mathcal{B}}$ generates $\tau_{[-\infty,+\infty]}$.
Q.E.D.

As we will show later on, the above result allows us to conclude that $\tau_{\mathbb{R}}$ and $\tau_{[-\infty,+\infty]}$ are second countable and thus define separable topological spaces.

1.3 Subspace Topologies

Let (E, τ) be a topological space. Sometimes we want to restrict the topology τ to a certain subset F of E. To this end, we can define the subspace topology on F induced by τ as

$$\tau_F = \{A \cap F \mid A \in \tau\}.$$

It is easy to check that τ_F is a well-defined topology on F; we call (F, τ_F) a subspace of (E, τ), and the elements of τ_F as sets that are open in F. Generally, a subset A of F is said to possess a topological property $\mathcal{P}$ in F if it possesses the property relative to the subspace topology τ_F. For instance, a set $A \subset F$ is closed in F if its complement $F \setminus A$ is open in F.

1.4 Metric Topologies

Let (E,d) be a metric space. Then, the collection

$$\mathbb{B} = \{B_d(x,\delta) \mid x \in E, \delta > 0\}$$

of open balls in E defines a base of E; to see this, note that, for any $x \in E$, $x \in B_d(x,1)$, so that

$$E \subset \bigcup_{x\in E} B_d(x,1) \subset \bigcup_{B\in\mathbb{B}} B.$$

In addition, for any $x,y \in E$ and $\delta_1,\delta_2 > 0$ such that $B_d(x,\delta_1) \cap B_d(y,\delta_2) \neq \emptyset$, choose any $z \in B_d(x,\delta_1) \cap B_d(y,\delta_2)$; then, $z \in B_d(z,\min(\delta_1 - d(x,z),\delta_2 - d(y,z))) \subset B_d(x,\delta_1) \cap B_d(y,\delta_2)$, where

$$\delta = \min(\delta_1 - d(x,z),\delta_2 - d(y,z)) > 0$$

because $d(x,z) < \delta_1$, $d(y,z) < \delta_2$, and $B_d(z,\delta) \in \mathbb{B}$. By definition, $\mathbb{B}$ is a base of E.
The metric topology on E induced by the metric d is the topology on E generated by the base $\mathbb{B}$ of open balls in E.

Metric topologies provide an alternative way to characterize open sets.

Lemma 1.6 Let τ be the metric topology on E induced by the metric d. Then, a nonempty $A \subset E$ is an open set if and only if, for any $x \in A$, there exists an $\varepsilon > 0$ such that $B_d(x,\varepsilon) \subset A$.

Proof) Suppose $A \in \tau$. Then, for any $x \in A$, because the base $\mathbb{B}$ of open balls generates τ, there exists a $B_d(z,\delta) \in \mathbb{B}$ such that $x \in B_d(z,\delta) \subset A$, where $z \in E$ and $\delta > 0$. Then, defining $\varepsilon = \delta - d(x,z) > 0$, for any $w \in B_d(x,\varepsilon)$

$$d(w,z) \leq d(w,x) + d(x,z) < \varepsilon + d(x,z) = \delta,$$

so that $B_d(x,\varepsilon) \subset B_d(z,\delta)$. As such,

$$B_d(x,\varepsilon) \subset B_d(z,\delta) \subset A;$$

this holds for any $x \in A$.

Conversely, suppose that, for any $x \in A$, there exists an $\varepsilon > 0$ such that $B_d(x,\varepsilon) \subset A$. Then, becasue $B_d(x,\varepsilon) \in \mathbb{B}$, this means that $x \in B_d(x,\varepsilon) \subset A$. This holds for any $x \in A$, A is in the topology generated by $\mathbb{B}$ by definition. However, τ is the topology generated by $\mathbb{B}$, so $A \in \tau$.
Q.E.D.

The euclidean topology on $\mathbb{R}^n$ is defined as the metric topology induced by the euclidean metric d on $\mathbb{R}^n$. A useful property of the euclidean topology on $\mathbb{R}$ is that it is exactly the standard topology on $\mathbb{R}$.

Theorem 1.7 The standard topology τ and the euclidean topology τ^e on $\mathbb{R}$ are equivalent.

Proof) Let $\mathbb{B}$ be the base of open balls under the euclidean metric, and $\mathbb{B}'$ the base consisting of open intervals in $\mathbb{R}$. Note that any open interval $(a,b) \in \mathbb{B}'$ can be seen as an open ball $(a,b) = B_d\left(\frac{a+b}{2}, \frac{b-a}{2}\right)$, so that $(a,b) \in \mathbb{B}$ and thus $\mathbb{B}' \subset \mathbb{B}$. Since $\mathbb{B}$ is contained in τ^e, so is $\mathbb{B}'$. By theorem 1.3, τ is the smallest topology containing $\mathbb{B}'$, so that $\tau \subset \tau^e$.

Likewise, any open ball $B_d(x,\delta) \in \mathbb{B}$ can be viewed as an open interval $(x-\delta, x+\delta) \in \mathbb{B}'$. As such, $\mathbb{B} \subset \mathbb{B}' \subset \tau$, and since τ^e is the smallest topology containing $\mathbb{B}$ and τ is a topology containing $\mathbb{B}$, $\tau^e \subset \tau$. It follows that $\tau = \tau^e$.
Q.E.D.

1.5 Product Topologies

Let there be n topological spaces $(E_1,\tau_1),\cdots,(E_n,\tau_n)$, and define the collection

$$\mathbb{B} = \{A_1 \times \cdots \times A_n \mid A_1 \in \tau_1, \cdots, A_n \in \tau_n\}$$

of open rectangles in $E = E_1 \times \cdots \times E_n$. This collection is a base on E; $E = E_1 \times \cdots \times E_n \subset \bigcup_{B\in\mathbb{B}} B$ because $E_1 \times \cdots \times E_n$ is an open rectangle. In addition, for any $A = A_1 \times \cdots \times A_n, B = B_1 \times \cdots \times B_n \in \mathbb{B}$ such that

$$A \cap B = (A_1 \cap B_1) \times \cdots \times (A_n \cap B_n) \neq \emptyset,$$

$A_i \cap B_i$ for $1 \leq i \leq n$ and each $A_i \cap B_i \in \tau_i$, so that, putting

$$C = (A_1 \cap B_1) \times \cdots \times (A_n \cap B_n),$$

$C \in \mathbb{B}$ and $C = A \cap B$. It follows that $\mathbb{B}$ is a base on E.

The product topology τ of the topologies $\tau_1,\cdots,\tau_n$ is the topology on E generated by the collection $\mathbb{B}$ of all open rectangles on E; it is often denoted as

$$\tau = \prod_{i=1}^{n} \tau_i = \tau_1 \times \cdots \times \tau_n.$$

The following result shows that the product of bases serves itself as a base of the product topology.

Lemma 1.8 Let $\mathbb{B}_i$ be a base generating τ_i for $1 \leq i \leq n$. Then, the collection $\mathbb{B}$ of subsets of $E = E_1 \times \cdots \times E_n$ defined as

$$\mathbb{B} = \{B_1 \times \cdots \times B_n \mid B_1 \in \mathbb{B}_1, \cdots, B_n \in \mathbb{B}_n\}$$

is a base on E generating $\tau = \tau_1 \times \cdots \times \tau_n$.

Proof) We show first that $\mathbb{B}$ is indeed a base of E.
Choose any $x = (x_1,\cdots,x_n) \in E$. Each $\mathbb{B}_i$ is a base of E_i, so for any $x_i \in E_i$, there exists a $B_i \in \mathbb{B}_i$ such that $x_i \in B_i$. As such, $x \in B_1 \times \cdots \times B_n \in \mathbb{B}$; this holds for any $x \in E$, so $E \subset \bigcup_{B\in\mathbb{B}} B$.
Now choose $A = A_1 \times \cdots \times A_n, B = B_1 \times \cdots \times B_n \in \mathbb{B}$ such that $A \cap B \neq \emptyset$, and let $x = (x_1,\cdots,x_n) \in A \cap B$. Then, $x_i \in A_i \cap B_i$ for $1 \leq i \leq n$, and by the property of bases, there exist $C_1 \in \mathbb{B}_1,\cdots,C_n \in \mathbb{B}_n$ such that $x_i \in C_i \subset A_i \cap B_i$ for $1 \leq i \leq n$. Therefore,

$$x \in C_1 \times \cdots \times C_n \subset A \cap B,$$

where $C_1 \times \cdots \times C_n \in \mathbb{B}$.

It remains to be seen that $\mathbb{B}$ generates τ. For any nonempty $A \in \tau$ and $x = (x_1, \cdots, x_n) \in A$, because τ is generated by the base collecting all open rectangles on E, there exist $A_1 \in \tau_1, \cdots, A_n \in \tau_n$ such that

$$x \in A_1 \times \cdots \times A_n \subset A.$$

For each $1 \leq i \leq n$, because τ_i is generated by $\mathbb{B}_i$, there exists a $B_i \in \mathbb{B}_i$ such that $x_i \in B_i \subset A_i$; it follows that

$$x \in B_1 \times \cdots \times B_n \subset A_1 \times \cdots \times A_n \subset A,$$

where $B_1 \times \cdots \times B_n \in \mathbb{B}$.
This holds for any element of a nonempty set in τ, and $\mathbb{B}$, being a subcollectin of open rectangles on E, is clearly contained in τ, so by lemma 1.2, $\mathbb{B}$ is a base generating τ.
Q.E.D.

In the case of an arbitrary collection of topological spaces $\{(E_i, \tau_i)\}$, the product topology $\tau = \prod_i \tau_i$ on the product space $E = \prod_i E_i$ is the topology generated by the base of open rectangles $\mathbb{B}$ defined as follows:

$$\mathbb{B} = \Big\{ \prod_i A_i \mid \forall i, A_i \in \tau_i, \text{ and finitely many of the } A_i \text{ are different from } E_i \Big\}.$$

It is again easy to show that this constitutes a base on E.

1.5.1 The Euclidean Topology on $\mathbb{R}^n$

An important result concerning product topologies is that the euclidean topology τ_n^e on $\mathbb{R}^n$ is the product of n standard topologies of $\mathbb{R}$, that is,

$$\tau_n^e = \underbrace{\tau_{\mathbb{R}} \times \cdots \times \tau_{\mathbb{R}}}_{n} = \tau_{\mathbb{R}}^n.$$

To show this, we define the square metric ρ on $\mathbb{R}^n$ as

$$\rho(x,y) = \max_{1 \le i \le n} |x_i - y_i|$$

for any $x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_n) \in \mathbb{R}^n$. ρ is indeed a metric on $\mathbb{R}^n$:

i) $\rho(x,x) = \max_{1 \le i \le n} |x_i - x_i| = 0$ for any $x \in \mathbb{R}^n$.

ii) If $x, y \in \mathbb{R}^n$ and $x \neq y$, then there exists a $1 \le j \le n$ such that $x_j \neq y_j$, so $\rho(x,y) = \max_{1 \le i \le n} |x_i - y_i| \ge |x_j - y_j| > 0$.

iii) For any $x, y \in \mathbb{R}^n$,

$$\rho(x,y) = \max_{1 \le i \le n} |x_i - y_i| = \max_{1 \le i \le n} |y_i - x_i| = \rho(y,x).$$

iv) For any $x, y, z \in \mathbb{R}^n$,

$$\begin{aligned} \rho(x,z) = \max_{1 \le i \le n} |x_i - z_i| &\le \max_{1 \le i \le n} [|x_i - y_i| + |y_i - z_i|] \\ &\le \max_{1 \le i \le n} |x_i - y_i| + \max_{1 \le i \le n} |y_i - z_i| = \rho(x,y) + \rho(y,z). \end{aligned}$$

Let d^n be the euclidean metric on $\mathbb{R}^n$ and d the eucliden metric on $\mathbb{R}$.

Theorem 1.9 Let the metric topology on $\mathbb{R}^n$ induced by d^n be τ_n^e, and that induced by ρ be denoted τ_n^s. Then,

$$\tau_n^e = \tau_n^s = \tau_{\mathbb{R}}^n.$$

Proof) Denote by $\mathbb{B}^e$, $\mathbb{B}^s$ and $\mathbb{B}^r$ the collection of open balls under d^n, the collection of open balls under ρ, and the collection of all open rectangles on $\mathbb{R}^n$, respectively. Note that $\mathbb{B}^e$, $\mathbb{B}^s$ and $\mathbb{B}^r$ are bases of E that generate the topologies τ_n^e, τ_n^s and $\tau_{\mathbb{R}}^n$ on E.

We first show that $\tau_n^e = \tau_n^s$. For any nonempty $A \in \tau_n^e$ and $x \in A$, it was seen in lemma 1.6 that there exists a $\delta > 0$ such that $B_{d^n}(x, \delta) \subset A$. For any $z \in B_\rho\left(x, \frac{\delta}{n}\right)$,

$$d^n(x,z) = \sqrt{\sum_{i=1}^n (z_i - x_i)^2} \le \sum_{i=1}^n |z_i - x_i| \le n \cdot \left[\max_{1 \le i \le n} |z_i - x_i|\right] = n\rho(x,z) < \delta,$$

so that $z \in B_{d^n}(x, \delta)$. Therefore, letting $\varepsilon = \frac{\delta}{n}$,

$$B_\rho(x, \varepsilon) \subset B_{d^n}(x, \delta) \subset A;$$

this shows us that, for any $x \in A$, there exists a $B \in \mathbb{B}^s$ centered at x such that $B \subset A$. By lemma 1.6, this indicates that $A \in \tau_n^s$, so that $\tau_n^e \subset \tau_n^s$.

To see the reverse inclusion, choose any nonempty $A \in \tau_n^s$. Then, for any $x \in A$, by lemma 1.6 there exists a $\delta > 0$ such that $B_\rho(x, \delta) \subset A$. For any $z \in B_{d^n}(x, \delta)$, letting $\rho(x, z) = max_{1 \leq i \leq n}|x_i - z_i| = |x_j - z_j|$ for some $1 \leq j \leq n$, we have

$$\rho(x, z) = |x_j - z_j| \leq \sqrt{\sum_{i=1}^{n}(x_i - z_i)^2} = d(x, z) < \delta,$$

so that $z \in B_\rho(x, \delta)$ as well; this implies that

$$B_{d^n}(x, \delta) \subset B_\rho(x, \delta) \subset A.$$

This holds for any $x \in A$, so by lemma 1.6, $A \in \tau_n^e$ and $\tau_n^s \subset \tau_n^e$.
Therefore, we have the relationship $\tau_n^e = \tau_n^s$.

Next, we show that $\tau_n^s = \tau_{\mathbb{R}}^n$. For any nonempty $A \in \tau_n^s$ and $x \in A$, there exists a $\delta > 0$ such that $B_\rho(x, \delta) \subset A$. Defining $A_i = (x_i - \delta, x_i + \delta)$ for $1 \leq i \leq n$, for any $z \in A_1 \times \cdots \times A_n$, $z_i \in (x_i - \delta, x_i + \delta)$ and thus $|x_i - z_i| < \delta$ for any $1 \leq i \leq n$, which implies that

$$\rho(z, x) = \max_{1 \leq i \leq n} |z_i - x_i| < \delta,$$

or $z \in B_\rho(x, \delta)$. As such,

$$x \in A_1 \times \cdots \times A_n \subset B_\rho(x, \delta) \subset A.$$

Since $A_1 \times \cdots \times A_n \in \mathbb{B}^r$, we have seen that, for any $x \in A$, there exists a $B \in \mathbb{B}^r$ such that $x \in B \subset A$. By definition of a topology generated by a base, this means that $A \in \tau_{\mathbb{R}}^n$; this holds for any nonempty $A \in \tau_n^s$, so $\tau_n^s \subset \tau_{\mathbb{R}}^n$.

Likewise, for any nonempty $A \in \tau_{\mathbb{R}}^n$ and $x \in A$, there exists an open rectangle $B = A_1 \times \cdots \times A_n \in \mathbb{B}^r$ such that $x \in B \subset A$. For $1 \leq i \leq n$, since $A_i \in \tau_{\mathbb{R}}$ and $x_i \in A_i$, by lemma 1.6 there exists a $\delta_i > 0$ such that $B_d(x_i, \delta_i) = (x_i - \delta_i, x_i + \delta_i) \subset A_i$. Letting $\delta = \min_{1 \leq i \leq n} \delta_i > 0$, for any $z \in B_\rho(x, \delta)$ and $1 \leq i \leq n$,

$$|z_i - x_i| \leq \max_{1 \leq j \leq n} |z_j - x_j| = \rho(z, x) < \delta,$$

so that

$$z_i \in (x_i - \delta, x_i + \delta) \subset (x_i - \delta_i, x_i + \delta_i) \subset A_i.$$

Therefore, $z \in A_1 \times \cdots \times A_n = B$, and

$$B_\rho(x, \delta) \subset B \subset A.$$

This shows us that, for any $x \in A$, there exists a $B \in \mathbb{B}^s$ centered at x such that $B \subset A$. By lemma 1.6, this indicates that $A \in \tau_n^s$, so that $\tau_{\mathbb{R}}^n \subset \tau_n^s$ and therefore $\tau_n^s = \tau_{\mathbb{R}}^n$.
It now follows that $\tau_n^e = \tau_n^s = \tau_{\mathbb{R}}^n$.
Q.E.D.

1.6 Second Countability and Separability

Let (E,τ) be a topological space. An open neighborhood of $x \in E$ under τ is any open set $A \in \tau$ such that $x \in A$.
We say that a set $A \subset E$ is dense in E under τ if, for any $x \in E$ and open neighborhood $N \in \tau$ of x, $A \cap N \neq \emptyset$; in other words, if an element of A is present in any collection of points arbitrarily close to x. Note that A need not be open.

(E,τ) is said to be second countable if there exists a countable base of E that generates τ.
On the other hand, (E,τ) is said to be separable if there exists a countable subset A of E that is dense in E under τ.

We can show that second countability implies separability. The converse does not hold true in general, but it is true for metrizable spaces, or topological spaces for which there exists a metric such that the associated topology is induced by that metric.

Theorem 1.10 The following hold true:

i) If (E,τ) is second countable, then it is separable.

ii) Suppose (E,τ) is metrizable. In this case, if (E,τ) is separable, then it is second countable.

Proof) We prove the claims in turn.
Suppose (E,τ) is second countable. Then, there exists a countable base $\mathbb{B}$ of E that generates τ; due to the countability of $\mathbb{B}$, we can arrange the elements of $\mathbb{B}$ into a sequence $\{B_n\}_{n \in N_+}$. Now construct the sequence $A = \{x_n\}_{n \in N_+}$ as follows: for any $n \in N_+$, if $B_n = \emptyset$, then let x_n be any point in E, while if $B \neq \emptyset$, choose x_n to be any point in B_n. Then, A is a countable subset of E (that is not necessarily open).
Next, choose any $x \in E$ and let $N \in \tau$ be any neighborhood of x. Because $\mathbb{B}$ generates τ and $x \in N$, by definition there exists a $B_i \in \mathbb{B}$ such that $x \in B_i \subset N$. Then,

$$x_i \in B_i \subset N \quad \text{and} \quad x_i \in A,$$

so that $\{x_i\} \subset N \cap A$ and thus $A \cap N \neq \emptyset$. This holds for any neighborhood of x and any point $x \in E$, so by definition A is dense in E under τ. Because A is countable, by definition (E,τ) is separable.

Now suppose (E,τ) is metrizable using the metric d on E, and that (E,τ) is separable. By definition, there exists a countable $A \subset E$ that is dense in E under τ; by countability, the elements of A can be arranged into a sequence $\{a_n\}_{n \in N_+}$. Define the collection

$$\mathbb{B} = \{B_d(a_n, q) \mid n \in N_+, q \in \mathbb{Q}\};$$

then, $\mathbb{B}$ is a countable collection of open balls (we can construct a one-to-one correspondence between $\mathbb{B}$ and the countable set $N_+ \times \mathbb{Q}$) and thus $\mathbb{B}$ is contained in τ (τ is the metric topology induced by d and is thus generated by the base of all open balls).
For any nonempty $B \in \tau$ and $x \in B$, by lemma 1.6 there exists a $\delta > 0$ such that $B_d(x,\delta) \subset B$. Because $B_d\left(x, \frac{\delta}{2}\right)$ is an open neighborhood of x and A is dense in E under τ, $B_d\left(x, \frac{\delta}{2}\right) \cap A \neq \emptyset$.
Letting $a_n \in B_d\left(x, \frac{\delta}{2}\right) \cap A$, we can see that $x \in B_d\left(a_n, \frac{\delta}{2}\right)$, and, by the denseness of $\mathbb{Q}$ in $\mathbb{R}$, we can choose a $q \in \mathbb{Q}$ such that

$$d(x,a_n) < q < \frac{\delta}{2},$$

so that $x \in B_d(a_n,q)$, where $B_d(a_n,q) \in \mathbb{B}$.
For any $z \in B_d(a_n,q)$,

$$d(z,x) \leq d(z,a_n) + d(a_n,x) < 2q < \delta,$$

so $z \in B_d(x,\delta)$ and

$$x \in B_d(a_n,q) \subset B_d(x,\delta) \subset B.$$

Thus, $\mathbb{B}$ is a collection of subsets of E contained in τ such that, for any nonempty $B \in \tau$ and $x \in B$, there exists a $B' \in \mathbb{B}$ such that $x \in B' \subset B$. By lemma 1.2, $\mathbb{B}$ is a base of E that generates τ. Since $\mathbb{B}$ was already seen to be countable, this proves that (E,τ) is second countable.
Q.E.D.

Second countability is an especially desirable property because it is preserved across Cartesian products.

Lemma 1.11 Let $(E_1,\tau_1),\cdots,(E_n,\tau_n)$ be topological spaces, and $\tau = \prod_{i=1}^n \tau_i$ the product of $\tau_1,\cdots,\tau_n$. If $\tau_1,\cdots,\tau_n$ are second countable with countable bases $\mathbb{B}_1,\cdots,\mathbb{B}_n$, then the countable collection of open rectangles

$$\mathbb{B} = \{A_1 \times \cdots \times A_n \mid A_1 \in \mathbb{B}_1,\cdots,A_n \in \mathbb{B}_n\}$$

is a base of τ. As a result, τ is also second countable.

Proof) This follows easily from lemma 1.8. Specifically, since $\mathbb{B}_1,\cdots,\mathbb{B}_n$ are bases of $E_1,\cdots,E_n$ generating $\tau_1,\cdots,\tau_n$, the collection $\mathbb{B}$ generates the product τ of $\tau_1,\cdots,\tau_n$.
Q.E.D.

1.7 Closures and Interiors of Sets

Let (E, τ) be a topological space. Recall that any set $A \subset E$ is said to be closed if $A^c \in \tau$, or A^c is open. From this, it follows that:

- E, $\emptyset$ are closed sets; $E^c = \emptyset \in \tau$, and likewise, $\emptyset^c = E \in \tau$.

- For any arbitrary collection $\{A_\alpha\}$ of closed subsets of E, because $A_\alpha^c \in \tau$ for each α, $\bigcup_\alpha A_\alpha^c \in \tau$ by the definition of a topology and thus
$$\bigcap_\alpha A_\alpha = \left(\bigcup_\alpha A_\alpha^c\right)^c$$
is closed.

- For any finite collection $\{A_1, \cdots, A_n\}$ of closed subsets of E, because $A_i^c \in \tau$ for $1 \le i \le n$, $\bigcap_{i=1}^n A_i^c \in \tau$ and thus
$$\bigcup_{i=1}^n A_i = \left(\bigcap_{i=1}^n A_i^c\right)^c$$
is also closed.

Note that closedness in subspace topologies has the same convenient characterization as openness. Let F be a subset of E, and τ_F its subspace topology. For any set A that is closed in F, note that $A = F \setminus V$ for some set V that is open in F, that is, contained in τ_F. By the definition of τ_F, there exists a $G \in \tau$ such that $V = G \cap F$, so we have
$$A = F \setminus (G \cap F) = F \cap (G^c \cup F^c) = F \cap G^c,$$
where G^c is closed relative to the whole topology τ. Therefore, any set that is closed in F can be represented as the intersection of F and a set closed relative to the whole topology.

For any $A \subset E$, then, A is contained in at least one closed set E. Therefore, the collection
$$\{B \subset E \mid A \subset B,\ B \text{ is closed}\}$$
is nonempty and therefore we can define the closure $\bar{A}$ of A as
$$\bar{A} = \bigcap_{A \subset B,\ B \text{ is closed}} B,$$
that is, as the intersection of all closed sets containing A. Since the intersection of closed sets is also closed, $\bar{A}$ can be viewed as the smallest closed set containing A.

Note that, if A is closed, then it is itself the smallest closed set containing A, and thus $\bar{A} = A$. On the other hand, if $\bar{A} = A$, then A is closed because $\bar{A}$ is. It follows that a necessary and sufficient condition for A to be closed is for $\bar{A} = A$, or for A to be equal to its closure.

A related concept is the interior of a set. Let $A \subset E$. Since A contains at least one open set, namely the empty set, it follows that the interior A^o of A is well-defined as

$$A^o = \bigcup_{B \subset A,\ B \text{ is open}} B,$$

that is, as the union of all open sets contained in A. Since the union of open sets is also open, A^o can be viewed as the largest open set contained in A.
Note that, if A is open, then it is itself the largest open set contained in A, and thus $A^o = A$. On the other hand, if $A^o = A$, then A is open because A^o is. It follows that a necessary and sufficient condition for A to be open is for $A^o = A$, or for A to be equal to its interior.

For any $x \in E$, we call any open set $N \in \tau$ containing x a neighborhood of x.
For any $A \subset E$ and $x \in E$, $x \in E$ a limit point of A if, for any neighborhood $N \in \tau$ of x, $A \cap (N \setminus \{x\}) \neq \emptyset$, that is, if we can find elements of A arbitrarily close to x. We denote the set of all limit points of A by A'.
Given the concept of limit points, we can define the limit of any sequence $\{x_n\}_{n \in N_+} \subset E$. We say that $x \in E$ is a limit of the sequence $\{x_n\}_{n \in N_+} \subset E$ if:

For any neighborhood N of x, there exists an $N \in N_+$ such that $x_n \in N$ for any $n \geq N$.

Note that this does not say anything about the uniqueness of the limit x or other familiar property of limits.

Armed with these concepts, we can furnish the following characterizations of the closure and interior of sets:

Lemma 1.12 Let (E, τ) be a topological space, and $A \subset E$. Then,

$$\bar{A} = A' \cup A = \{x \in E \mid A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x\},$$

and

$$A^o = \{x \in E \mid \exists N \in \tau, N \subset A \text{ such that } x \in N\}.$$

Proof) We will first show that $\bar{A} = A' \cup A$.
Choose any $x \in \bar{A}$. Suppose that $x \notin A$ and that there exists some neighborhood $N \in \tau$ of x such that $A \cap (N \setminus \{x\}) = A \cap N = \emptyset$, where the first equality follows from the assumption that $x \notin A$. Then, N^c is a closed set containing A, and $\bar{A} \subset N^c$ by definition.

Since $x \in \bar{A}$, this means that $x \notin N^c$, a contradiction.
Therefore, every neighborhood $N \in \tau$ of x must satisfy $A \cap (N \setminus \{x\}) \neq \emptyset$, and by definition $x \in A'$.
x is either contained in A or a limit point of A, so $x \in A \cup A'$. This holds for any $x \in \bar{A}$, so $\bar{A} \subset A' \cup A$.

To see the converse, let B be a closed set containing A. For any $x \in A'$, suppose $x \notin B$. This means that $x \in B^c \in \tau$, so B^c is a neighborhood containing x; by the definition of a limit point, there exists a $y \in A \cap (B^c \setminus \{x\})$, so that $y \in A$ and $y \in B^c$. This is a contradiction because $y \in A$ and $A \subset B$ necessarily imply that $y \in B$.
Therefore, $x \in B$, meaning that $A' \cup A \subset B$. $\bar{A}$ is a closed set contaning A, so it follows that $A' \cup A \subset \bar{A}$, and finally that $\bar{A} = A' \cup A$.

It is easy to see that $A' \cup A = \{x \in E \mid A \cap N \neq \emptyset \text{ for any neighborhood } N \text{ of } x\}$. Let $x \in A' \cup A$; if $x \in A$, then for any neighborhood $N \in \tau$ of x, $x \in A \cap N$. On the other hand, if $x \notin A$, then for any neighborhood $N \in \tau$ of x, by definition $A \cap N = A \cap (N \setminus \{x\}) \neq \emptyset$. Conversely, for any $x \in E$ such that $A \cap N \neq \emptyset$ for any neighborhood N of x, either $x \in A$ or, if $x \notin A$, then $A \cap (N \setminus \{x\}) = A \cap N \neq \emptyset$ for any neighborhood N of x, so that $x \in A'$.

Finally, we prove the characterization of A^{o}.
For any $A \subset E$, let $x \in A^{\mathrm{o}}$. Then, A^{o} is itself an open set containing x and contained in A, so that x is an element of the set $\{y \in E \mid \exists N \in \tau, N \subset A \text{ such that } y \in N\}$.
Conversely, suppose that, for some $x \in E$, there exists a neighborhood N of x such that $N \subset A$. Then, N is an open set contained in A, meaning that $N \subset A^{\mathrm{o}}$ by definition; it follows that $x \in A^{\mathrm{o}}$.
Q.E.D.

1.8 Compact Sets and Hausdorff Spaces

Let (E, τ) be a topological space. We say that some subset A of E is compact if, for any open cover $\{V_\alpha\}$ of A, that is, a collection of open sets in E such that $A \subset \bigcup_\alpha V_\alpha$, there exists a finite subset $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ of $\{V_\alpha\}$ such that

$$A \subset V_{\alpha_1} \cup \cdots \cup V_{\alpha_n}.$$

The collection $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ is called a subcover of A; therefore, A is compact if any open cover of A has a finite subcover.
We say that (E, τ) is a compact space if the entire set E is itself compact.

A related concept is local compactness. (E, τ) is a locally compact topological space if any $x \in E$ has a neighborhood $N \in \tau$ such that $\bar{N}$ is compact.
Note that any compact space is also locally compact. To see this, let (E, τ) be a compact space and choose any $x \in E$; then, E is a neighborhood of x whose closure is $\bar{E} = E$ because E is closed, and E is compact by definition.
In addition, recall the Heine-Borel theorem, which posits that every closed and bounded subset of a euclidean space $\mathbb{R}^n$ is compact. As such, for any $x \in \mathbb{R}^n$, x is contained in an open set $(-|x|, |x|)^n$, which is contained in the closed and bounded and thus compact set $[-|x|, |x|]^n$. This tells us that any euclidean topological space is locally compact.

A topological space (E, τ) is said to be Hausdorff if, for any distinct points $x, y \in E$, there exist neighborhoods $N_x, N_y \in \tau$ of x, y such that $N_x \cap N_y = \emptyset$. In other words, any two distinct points in a Hausdorff space can be separated by open sets. This is called the separation, or T^2, property.
Every metric space (E, d) is Hausdorff under the metric topology τ induced by d. For any $x, y \in E$ such that $x \neq y$, defining $\varepsilon = d(x, y) > 0$, the open balls $B_d(x, \varepsilon/2)$ and $B_d(y, \varepsilon/2)$ are open sets containing x, y. If $z \in B_d(x, \varepsilon/2) \cap B_d(y, \varepsilon/2)$, then

$$d(x, y) \leq d(z, x) + d(z, y) < \varepsilon,$$

a contradiction. Therefore, $B_d(x, \varepsilon/2) \cap B_d(y, \varepsilon/2) = \emptyset$, showing us that (E, τ) has the separation property.

Limits of sequences in Hausdorff spaces are unique; to see this, let $\{x_n\}_{n \in N_+}$ be a sequence in the Hausdorff space (E, τ), and let $x, y \in E$ be limits of $\{x_n\}_{n \in N_+}$. Suppose $x \neq y$. Then, there exist neighborhoods $N_x, N_y \in \tau$ of x, y such that $N_x \cap N_y = \emptyset$ by the separation property. By the definition of limits, there exist $m_1, m_2 \in N_+$ such that $x_n \in N_x$ for any $n \geq m_1$ and $x_n \in N_y$ for any $n \geq m_2$. Then, for any $n \geq \max(m_1, m_2)$, $x_n \in N_x \cap N_y = \emptyset$, a contradiction. Therefore, $x = y$ and the limit of sequences are unique.

1.8.1 Properties of Compact Sets

The following are properties of compact sets:

Theorem 1.13 Let (E,τ) be a topological space. The following hold true:

i) **(Preservation of Compactness across Subspaces)**
Let K,F be chosen so that $K \subset F \subset E$. K is compact in the subspace topology τ_F if and only if it is compact in the whole topology τ.

ii) If a subset K of E is compact and $F \subset K$ is a closed subset of E, then F is also compact.

iii) **(Finite Intersection Characterization of Compactness)**
A subset K of E is compact if and only if, for any collection $\{A_\alpha\}$ of sets that are closed in K such that every finite subcollection of $\{A_\alpha\}$ has a non-empty intersection, $\cap_\alpha A_\alpha$ is also non-empty.

iv) **(Second Countability of Compact Metric Spaces)**
If (E,τ) is metrizable, then it is second countable, that is, there exists a countable base on E that generates τ.

v) **(Regularity of Compact Hausdorff Spaces)**
Let (E,τ) be a Hausdorff space. If $K \subset E$ is compact and $x \notin K$ for some $x \in E$, then there exist open sets $A,B \in \tau$ such that $K \subset A$ and $x \in B$ such that $A \cap B = \emptyset$.

vi) Any compact subset K in a Hausdorff space (E,τ) is closed.

vii) For any compact subset K and closed subset F in a Hausdorff space (E,τ), $F \cap K$ is compact.

viii) Let (E,τ) be a Hausdorff space. If $\{K_\alpha\}$ is a collection of compact subsets of E with $\cap_\alpha K_\alpha = \emptyset$, then there exists a finite subcollection of $\{K_\alpha\}$ whose intersection is empty.

Proof) i) Suppose that K is compact in the entire topology τ. We must show that K is also compact in the subspace topology τ_F. To this end, choose any cover $\{V_\alpha\}$ of K that is open in the subspace topology τ_F; by the definition of the subspace topology, for any α there exist $G_\alpha \in \tau$ such that $V_\alpha = G_\alpha \cap F$. This means that $\{G_\alpha\}$ forms a cover of K that is open in the entire topology τ, and by the compactness of K, there exist $\alpha_1, \cdots, \alpha_n$ such that

$$K \subset G_{\alpha_1} \cup \cdots \cup G_{\alpha_n}.$$

As such,

$$K = K \cap F \subset (G_{\alpha_1} \cup \cdots \cup G_{\alpha_n}) \cap F = \bigcup_{i=1}^{n} V_{\alpha_i},$$

and $\{V_\alpha\}$ has a finite subcover of K. This shows us that K is compact in the subspace topology τ_F.

Conversely, suppose that K is compact in the subspace topology τ_F. We must show that K is also compact in the entire topology τ, and to this end, choose an open cover $\{V_\alpha\} \subset \tau$ of K. Defining $G_\alpha = V_\alpha \cap F$ for any α, we can see that $\{G_\alpha\} \subset \tau_F$ that covers K, since K is contained in F. As such, by the compactness of K in τ_F, there exist $\alpha_1, \cdots, \alpha_n$ such that

$$K \subset G_{\alpha_1} \cup \cdots \cup G_{\alpha_n} = (V_{\alpha_1} \cup \cdots \cup V_{\alpha_n}) \cap F.$$

Therefore, $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ covers K, which shows us that K is compact in τ.

ii) Suppose K is compact and let F be a closed subset of K. For any open cover $\{V_\alpha\}$ of F, note that

$$K \subset \left(\bigcup_\alpha V_\alpha \right) \cup F^c,$$

since every element of K is either in F or not in F. Since $F^c \in \tau$ is open by definition, it follows that $\{\{V_\alpha\}, F^c\}$ is an open cover of K, and by compactness, there exists a finite subcover $\{V_{\alpha_1}, \cdots, V_{\alpha_n}, F^c\}$ of K. Since

$$F \subset K \subset \left(\bigcup_{i=1}^n V_{\alpha_i} \right) \cup F^c,$$

and $F \cap F^c = \emptyset$, we have

$$F \subset \bigcup_{i=1}^n V_{\alpha_i}.$$

Therefore, any open cover of F has a finite subcover, meaning that F is compact.

iii) Let K be a compact subset of E, and choose any collection $\{A_\alpha\}$ of subsets of K that are closed in K with non-empty finite intersections. We must show that $\cap_\alpha A_\alpha \neq \emptyset$. Suppose that this intersection is actually empty. Fixing any index α_1, note that A_{α_1}, being a closed subset of the comapct set K, is itself a compact set. Furthermore, by the first result, this implies that A_{α_1} is also compact in K.
For any $x \in A_{\alpha_1}$, there exists an α_x such that $x \notin A_{\alpha_x}$, since otherwise x belongs to the intersection $\cap_\alpha A_\alpha$. This shows us that, in terms of the subspace topology of K, $\{A^c_{\alpha_x}\}_{x \in A_{\alpha_1}}$ is an open cover of the compact set A_{α_1}. By definition, there

exist $x_1, \cdots, x_n \in A_{\alpha_1}$ such that

$$A_{\alpha_1} \subset \bigcup_{i=1}^{n} A^c_{\alpha_{x_i}}.$$

By implication,

$$A_{\alpha_1} \cap A_{\alpha_{x_1}} \cap \cdots \cap A_{\alpha_{x_n}} = \emptyset,$$

which contradicts the fact that $\{A_\alpha\}$ has non-empty finite intersections. As such, $\cap_\alpha A_\alpha$ must be non-empty.

Conversely, suppose that, for any collection $\{A_\alpha\}$ of closed subsets of K that have non-empty finite intersections, $\cap_\alpha A_\alpha \neq \emptyset$. We must show that K is compact. As above, suppose the contrary, that K is not compact; then, there exists an open cover $\{V_\alpha\}$ of K that has no finite subcover. Then, defining

$$G_\alpha = K \cap V^c_\alpha$$

for any α, $\{G_\alpha\}$ is a collection of sets closed in K (each V^c_α is closed relative to the whole topology) with non-empty finite intersections; if there exists a finite collection $\alpha_1, \cdots, \alpha_n$ such that

$$G_{\alpha_1} \cap \cdots \cap G_{\alpha_n} = K \cap \left(\bigcup_{i=1}^{n} V_{\alpha_i} \right)^c$$

is empty, then $\{V_{\alpha_1}, \cdots, V_{\alpha_n}\}$ forms a finite subcover of K, a contradiction. By assumption, $\bigcap_\alpha G_\alpha \neq \emptyset$. Choosing any $x \in \bigcap_\alpha G_\alpha$, we can see that $x \in K$ but $x \notin \cup_\alpha V_\alpha$, which contradicts the fact that K is covered by $\{V_\alpha\}$. Therefore, it must be the case that K is compact.

iv) Let (E, d) be a compact metric space and τ the metric topology induced by d. For any $n \in N_+$, $\{B_d(x, 1/n)\}_{x \in E}$ forms an open cover of E, and by the compactness of E, there exist a finite number of points $x^{(n)}_1, \cdots, x^{(n)}_{m_n} \in E$ such that

$$E \subset \bigcup_{i=1}^{m_n} B_d(x^{(n)}_i, 1/n).$$

Defining

$$\mathbb{B} = \{B_d(x^{(n)}_i, 1/n) \mid n \in N_+, \ 1 \leq i \leq m_n\},$$

$\mathbb{B}$ is a countable collection of open subsets of E that trivially covers E. We will show that $\mathbb{B}$ is a countable base of E that generates the metric topology τ.

For any $A \in \tau$ and $x \in A$, there exists an $\epsilon > 0$ such that $B_d(x, \epsilon) \subset A$. Letting $n \in N_+$ be chosen so that $\frac{1}{n} < \frac{\epsilon}{2}$, since E is covered by the finite collection $\{B_d(x_i^{(n)}, 1/n)\}_{1 \leq i \leq m_n}$, there exists some $1 \leq i \leq m_n$ such that

$$x \in B_d(x_i^{(n)}, 1/n).$$

Choose any $z \in B_d(x_i^{(n)}, 1/n)$. Then,

$$d(x,z) \leq d(x, x_i^{(n)}) + d(x_i^{(n)}, z) < \frac{2}{n} < \epsilon,$$

so that $z \in B_d(x, \epsilon) \subset A$. This shows us that

$$x \in B_d(x_i^{(n)}, 1/n) \subset A,$$

and since $B_d(x_i^{(n)}, 1/n)$ is an element of $\mathbb{B}$, we can immediately see that $\mathbb{B}$ is a base on E that generates τ.

v) Let (E, τ) be a Hausdorff space and assume K is a compact subset of E and $x \in E$ is not contained in K. If $K = \emptyset$, then E is a neighborhood of x and $\emptyset$ an open set containing K such that $\emptyset \cap E = \emptyset$, so the proof is complete.

On the other hand, let $K \neq \emptyset$. For any $y \in K$, because $x \neq y$ (otherwise, $x \in K$, a contradiction), by the separation property there exist neighborhoods $A_y, B_y \in \tau$ of x, y such that $A_y \cap B_y = \emptyset$.

Then, $\{B_y\}_{y \in K}$ is an open cover of K, and by compactness there exist $y_1, \cdots, y_n \in K$ such that

$$K \subset B_{y_1} \cup \cdots \cup B_{y_n} = B.$$

B is thus an open set containing K.

Define $A = A_{y_1} \cap \cdots \cap A_{y_n}$; it is the finite intersection of open sets and thus open itself. In addition, because each A_y is a neighborhood of x, A is also a neighborhood of x. Finally, if $z \in A \cap B$, then letting $z \in B_{y_i}$ for some $1 \leq i \leq n$, $z \in A_{y_i}$ as well, which contradicts the fact that $A_{y_i} \cap B_{y_i} = \emptyset$. Therefore, $A \cap B = \emptyset$, which concludes the proof.

vi) Let K be a compact subset in some Hausdorff space (E, τ), and let $x \in \bar{K}$ but $x \notin K$. By the preceding result, there exist open sets $A, B \in \tau$ such that $x \in A$, $K \subset B$ and $A \cap B = \emptyset$. By the characterization of the closure we derived earlier,

$A \cap K \neq \emptyset$, since $x \in \bar{K}$ and A is a neighborhood of x; however, because $K \subset B$ and $A \cap B = \emptyset$, $A \cap K = \emptyset$, a contradiction. Therefore, $x \in K$, and it follows that $\bar{K} \subset K$.

It follows by definition that $K \subset \bar{K}$, so $K = \bar{K}$. This shows us that K must be a closed set.

vii) Let K and F be compact and closed subsets, respectively, in a Hausdorff space (E, τ). Since K is closed by the preceding result, $F \cap K$ is a closed subset of the compact set K. It then follows from the very first result that $F \cap K$ is also compact.

viii) Let (E, τ) be a Hausdorff space, and $\{K_\alpha\}$ a collection of compact subsets of E whose intersection is empty. Choose some $K_1 \in \{K_\alpha\}$; for any $x \in K_1$, because $\bigcap_\alpha K_\alpha = \emptyset$, there exists an α such that $x \notin K_\alpha$ (otherwise, $x \in \bigcap_\alpha K_\alpha$, a contradiction). Labeling the complement of that set A_x, it follows that $x \in A_x$, and that $A_x \in \tau$, since compact sets in a Hausdorff space are closed.

Now $\{A_x\}_{x \in K_1}$ is an open cover of K_1, and by compactness, there exist $x_1, \cdots, x_n \in K_1$ such that

$$K_1 \subset A_{x_1} \cup \cdots \cup A_{x_n}.$$

Then, letting the complement of A_{x_i} be K_{x_i} for each $1 \leq i \leq n$, since each $K_{x_i} \in \{K_\alpha\}$ and

$$K_{x_1} \cap \cdots \cap K_{x_n} = (A_{x_1} \cup \cdots \cup A_{x_n})^c \subset K_1^c,$$

the finite intersection

$$K_1 \cap K_{x_1} \cap \cdots \cap K_{x_n}$$

is empty.

Q.E.D.

We now elaborate on some of the properties enumerated above. The first results shows us that compactness is a topological property that is preserved across subspaces. This is important because, in general, topological properties are not preserved across subspaces; for instance, given a set $F \subset E$, a subset of F open in F is generally not open in E unless F is an open subset of E. As such, when we talk about compactness, we may discuss compact spaces instead of compact subsets of a larger set, since, if we are given any compact subset of a topological space, the subspace induced by that compact subset becomes a compact space.

The third result is a useful characterization of compactness via the finite intersection property (FIP). A collection of sets $\{A_\alpha\}$ is said to possess the FIP if any finite intersection is non-empty; the characterization above states that K is compact if and only if any collection of sets that is closed in K and has the FIP has a non-empty intersection. We may succinctly state this as the characterization of K via the non-empty intersection property (NIP).

The fourth property states that any compact metric space is second countable. By the result on second countability and separability shown in a previous section, we can also see that any compact metric space is separable. This property is exploited in analysis to derive results on equicontinuous functions.

The fifth property is labeled here as the regularity of compact Hausdorff spaces. The concept of regular and normal spaces will be studied in more depth below.

We now derive a core property of locally compact Hausdorff spaces.

Theorem 1.14 Let (E,τ) be a locally compact Hausdorff space. Then, for any compact subset K of E and an open set $U \in \tau$ such that $K \subset U$, there exists an open set $V \in \tau$ with compact closure such that

$$K \subset V \subset \bar{V} \subset U.$$

Proof) Let K and U be compact and open subsets of the locally compact Hausdorff space (E,τ) such that $K \subset U$. For any $x \in K$, by the local compactness of (E,τ), there exists a neighborhood $B_x \in \tau$ of x with compact closure. $\{B_x\}_{x\in K}$ then forms an open cover of K, and by compactness, there exist $x_1,\cdots,x_n \in K$ such that

$$K \subset B_{x_1} \cup \cdots \cup B_{x_n} = G.$$

G is contained in the closed set $\bar{B_{x_1}} \cup \cdots \bar{B_{x_n}}$, so by definition $\bar{G} \subset \bar{B_{x_1}} \cup \cdots \bar{B_{x_n}}$. Each $\bar{B_{x_i}}$ is compact, and because the finite union of compact sets is compact (for any open cover of the union, there exists a finite subcover for each set comprising the union, and because the union is finite, the union of those finite subcovers is also finite), $\bar{G}$ is a closed subset of a compact set. It follows from the previous result that $\bar{G}$ is also compact.

G is an open set containing K with compact closure. Now we find a subset of G that contains K and whose closure is contained in U.

For any $x \in U^c$, because $K \subset U$ implies that $x \notin K$ and (E,τ) is Hausdorff, by the result proved earlier there exist open sets $A_x, C_x \in \tau$ such that $x \in C_x$, $K \subset A_x$ and $A_x \cap C_x = \emptyset$.

Suppose that $x \in \bar{A_x}$. Since $x \notin A_x$, x must be a limit point of A_x in this case. Becuase C_x is a neighborhood of x, this implies that $A_x \cap C_x = A_x \cap (C_x \setminus \{x\}) \neq \emptyset$, which

contradicts the fact that $A_x \cap C_x = \emptyset$. Therefore, $x \notin \bar{A}_x$, so that A_x is an open set containing K and whose closure does not contain x.
U^c is a closed subset of E because U is open, and $\bar{G}$ is compact, as shown above. $U^c \cap \bar{G} \cap \bar{A}_x$ is a closed subset of the compact set $\bar{G}$, from which it follows that $U^c \cap \bar{G} \cap \bar{A}_x$ is itself compact.

$\{U^c \cap \bar{G} \cap \bar{A}_x\}_{x \in U^c}$ is thus a collection of compact subsets in E. Suppose that the collection has a nonempty intersection, and denote by $z \in E$ an element of that intersection. Then, $z \in U^c$ and $z \in \bar{A}_z$ as well, which contradicts the fact that $z \notin \bar{A}_z$ by design. Therefore, $\{U^c \cap \bar{G} \cap \bar{A}_x\}_{x \in U^c}$ has an empty intersection, which implies, by lemma 4.2, that there exist $y_1, \cdots, y_m \in U^c$ such that

$$\bigcap_{i=1}^{m} \left(U^c \cap \bar{G} \cap \bar{A}_{y_i} \right) = U^c \cap \bar{G} \cap \bar{A}_{y_1} \cap \cdots \cap \bar{A}_{y_m} = \emptyset.$$

Defining

$$V = G \cap A_{y_1} \cap \cdots \cap A_{y_m} \in \tau,$$

since $K \subset G$ and $K \subset A_{y_i}$ for $1 \leq i \leq m$, V is an open set containing K. Furthermore, because

$$\bar{V} \subset \bar{G} \cap \bar{A}_{y_1} \cap \cdots \cap \bar{A}_{y_m} \subset U,$$

where the last inclusion follows from the fact that $U^c \cap \bar{G} \cap \bar{A}_{y_1} \cap \cdots \cap \bar{A}_{y_m} = \emptyset$, V is an open set such that

$$K \subset V \subset \bar{V} \subset U,$$

which is exactly the result we were looking for.
Finally, to show that $\bar{V}$ is compact, note that it is a closed subset of the compact set $\bar{G}$. From lemma 4.2, we can conclude that $\bar{V}$ is compact.
Q.E.D.

1.8.2 Notions of Compactness in Metric Spaces

Compact sets in metric spaces possess convenient properties. To illustrate these, we introduce some concepts related to compactness.

Let (E,τ) be a topological space and A a subset of E. A is

- **Limit Point Compact**
  If any infinite set in A has a limit point in A

- **Countably Compact**
  If any countable open cover of A has a finite subcover

- **Sequentially Compact**
  If any sequence in A has a convergent subsequence with limit in A

- **Relatively Compact**
  If the closure $\overline{A}$ of A is compact.

Clearly, a compact subset of a Hausdorff space is relatively compact (because it is closed and thus is equivalent to its closure) and any compact set is countably compact.

A related concept is that of a **Lindelöf Space**. A topological space (E,τ) is Lindelöf if any open cover of E has a countable subcover. Note how any compact space is also Lindelöf. We state below some relevant results.

Theorem 1.15 The following hold true:

i) Compact sets are limit point compact.

ii) Limit point compact sets in a metric space are sequentially compact.

iii) Metrizable sequentially compact spaces are second countable.

iv) Second countable topological spaces are Lindelöf.

v) Sequentially compact sets are countably compact.

vi) A subset of a metric space is compact if and only if it is sequentially compact.

Proof) i) Let K be a compact subset in the toplogical space (E,τ). Choose any infinite set A contained in K, and suppose that it does not have a limit point in K. Then, for any $x \in K$, since x is not a limit point of A, there exists a neighborhood V_x of x such that $(V_x \setminus \{x\}) \cap A = \emptyset$. The collection $\{V_x\}_{x\in K}$ forms an open cover of K, and by the compactness of K, there exist $x_1,\cdots,x_n \in K$ such that

$$K \subset \bigcup_{i=1}^{n} V_{x_i}.$$

Since $A \subset K$ and A is an infinite set, there exists an element $y \in A$ such that $y \neq x_i$ for any $1 \leq i \leq n$. It follows that $y \in V_{x_i}$ for some $1 \leq i \leq n$, and since $y \neq x_i$,

$$y \in V_{x_i} \cap A = (V_{x_i} \setminus \{x_i\}) \cap A.$$

This contradicts the fact that $(V_{x_i} \setminus \{x_i\}) \cap A = \emptyset$ by design, so A must have a limit point in K, meaning K is limit point compact.

ii) Let (E,d) be a metric space and K a limit point compact subset of E. Choose any sequence $\{x_n\}_{n \in N_+}$ in K. If $\{x_n\}_{n \in N_+}$ contains a finite number of distinct elements, then it trivially contains a convergent subsequence with limit in K, since there must be infinitely many elements of the sequence that assume the same value.

On the other hand, if $A = \{x_n \mid n \in N_+\}$ is an infinite set, then by the limit point compactness of K (which follows from i)), A has a limit point x in K. We can choose an $n_1 \in N_+$ such that $d(x, x_{n_1}) < 1$. Suppose that we have chosen integers $n_1 < \cdots < n_k$ for some $k \geq 1$. There then exists an $n_{k+1} > n_k$ such that

$$d(x, x_{n_{k+1}}) < \frac{1}{k+1},$$

where n_{k+1} can always be chosen to be larger than n_k since the neighborhood of x must contain infinitely many points in A. It follows that $\{x_{n_k}\}_{k \in N_+}$ is a subsequence of $\{x_n\}_{n \in N_+}$ that converges to x, proving the sequential compactness of K.

iii) Let (E,d) be a sequentially compact metric space. We will show that (E,d) is separable; since separability and second countability is equivalent for metric spaces, the second countability of (E,τ) then follows immediately.

Choose any $n \in N_+$ and $x_1^{(n)} \in E$. Assuming that $x_1^{(n)}, \cdots, x_k^{(n)} \in E$ have been chosen for some $k \geq 1$, choose $x_{k+1}^{(n)}$ as a point in E such that

$$d(x_i^{(n)}, x_{k+1}^{(n)}) \geq \frac{1}{n}$$

for any $1 \leq i \leq k$ if possible. Suppose that we are able to choose an infinite sequence $\{x_i^{(n)}\}_{i \in N_+}$ in this manner. Since E is sequentially compact, there exists a subsequence of $\{x_i^{(n)}\}_{i \in N_+}$ that converges to some point x in E; for notational brevity, suppose that the sequence $\{x_i^{(n)}\}_{i \in N_+}$ itself converges to x. Then, there exists some $N \in N_+$ such that

$$d(x, x_k^{(n)}) < \frac{1}{2n}$$

for any $k \geq N$, which implies that

$$d(x_N^{(n)}, x_{N+1}^{(n)}) < d(x, x_N^{(n)}) + d(x, x_{N+1}^{(n)}) < \frac{1}{n}.$$

However, this contradicts the fact that

$$d(x_i^{(n)}, x_{N+1}^{(n)}) \geq \frac{1}{n}$$

for any $1 \leq i \leq N$. Therefore, we must only be able to choose a finite number of points $x_i^{(n)}$ in the above manner; that is, there must exist some $m_n \in N_+$ such that, for any $x \in E$,

$$d(x_i^{(n)}, x) < \frac{1}{n}$$

for any $1 \leq i \leq m_n$.

Define the countable set A as

$$A = \{x_i^{(n)} \mid 1 \leq i \leq m_n, n \in N_+\}.$$

We now show that A is dense in E, and thus that E is separable. Choose any $x \in E$ and $\delta > 0$. For any $n \in N_+$ such that $\frac{1}{n} < \delta$, note that, from the way that we chose $x_1^{(n)}, \cdots, x_{m_n}^{(n)}$, there must exist some $1 \leq i \leq m_n$ such that $d(x_i^{(n)}, x) < \frac{1}{n} < \delta$, which implies that

$$x_i^{(n)} \in B_d(x, \delta) \cap A$$

and thus that $B_d(x, \delta) \cap A \neq \emptyset$. This holds for any $\delta > 0$, so it follows that $x \in \overline{A}$. This in turn holds for any $x \in E$, so $E \subset \overline{A}$. The reverse inclusion is trivial, and we can conclude that $E = \overline{A}$.

iv) Let (E, τ) be a second countable space with countable base $\mathbb{B} \subset \tau$. By the countability of $\mathbb{B}$, its elements can be arranged into a sequence $\mathbb{B} = \{B_n\}_{n \in N_+}$. Choose any open cover $\{V_\alpha\}$ of E. For any $x \in E$, there exists some α_x such that $x \in V_{\alpha_x}$, and since V_{α_x} is an open subset of E, there exists some $n_x \in N_+$ such that

$$x \in B_{n_x} \subset V_{\alpha_x}.$$

This shows us that E is covered by the collection $\{B_{n_x}\}_{x \in E}$, and by extension the subcollection $\{V_{\alpha_x}\}_{x \in E}$ of $\{V_\alpha\}$. Since $\{B_{n_x}\}_{x \in E}$ is a subset of $\mathbb{B}$, it is a countable collection, which makes $\{V_{\alpha_x}\}_{x \in E}$ countable as well. Therefore, E is covered by a countable subcollection of $\{V_\alpha\}$, making it a Lindelöf space.

v) Let (E, τ) be a topological space and K a sequentially compact subset of E. Choose a countable open cover $\{V_n\}_{n \in N_+}$ of K, and suppose that no finite collection of sets in $\{V_n\}_{n \in N_+}$ can cover K. Define the sequence $\{F_n\}_{n \in N_+}$ of subsets of E as

$$F_n = (V_1 \cup \cdots \cup V_n)^c.$$

Clearly, $\{F_n\}_{n \in N_+}$ is a sequence of decreasing closed sets, and by assumption, each F_n contains a point in K. Choose $x_n \in F_n$ for any $n \in N_+$; then, $\{x_n\}_{n \in N_+}$ is a sequence in K, and by the sequential compactness of K, there exists a subsequence $\{x_{n_k}\}_{k \in N_+}$ that converges to a point x in K. $\{V_n\}_{n \in N_+}$ covers K, so there exists some $n \in N_+$ such that $x \in V_n$, and by the definition of a limit, there also exists an $N \in N_+$ such that $x_{n_k} \in V_n$ for any $k \geq N$. This means that $x_{n_k} \notin F_n$, and since there must exist a $k \geq N$ such that $n_k > n$, we can see that $x_{n_k} \in F_{n_k}$ but $x_{n_k} \notin F_n$ for this k. However, the first inclusion indicates, together with $F_{n_k} \subset F_n$, that $x_{n_k} \in F_n$, a contradiction. Therefore, there must exist a finite subcollection of $\{V_n\}_{n \in N_+}$ that covers K, and by definition, K is countably compact.

vi) This result now follows from those proven above. Let (E, d) be a metric space and let K be a subset of E. Suppose initially that K is compact. Then, by i) and ii), it is limit point compact and thus sequentially compact.

Conversely, suppose that K is sequentially compact. Then, by iii) and iv), the space (K, d) is Lindelöf, meaning that any open cover $\{V_\alpha\}$ of K contains a countable subcover. Finally, since K is also countably compact by v), this countable subcover contains a finite subcover of K. This proves that K is compact.

Q.E.D.

Corollary to Theorem 1.15 (Bolzano-Weierstrass Theorem)
Let $F = \mathbb{R}^n$ or $\mathbb{C}$. Then, any bounded sequence $\{x_n\}_{n \in N_+}$ in F contains a convergent subsequence.

Proof) First let $F = \mathbb{R}^n$. Since $\{x_n\}_{n \in N_+}$ is bounded, there exists an $M < +\infty$ such that $|x_n| \leq M$ for any $n \in N_+$. This tells us that $\{x_n\}_{n \in N_+}$ is contained in the n-cell $[-M, M]^n \subset F$. Since n-cells are known to be compact, $\{x_n\}_{n \in N_+}$ is contained in a

compact subset of F; theorem 2.6 then implies that $\{x_n\}_{n\in N_+}$ contains a convergent subsequence.
Now let $F = \mathbb{C}$. Then, $\{Re(x_n)\}_{n\in N_+}$ is a bounded sequence in $\mathbb{R}$, and by the preceding result, there exists a convergent subsequence $\{Re(x_{n_k})\}_{k\in N_+}$ of $\{Re(x_n)\}_{n\in N_+}$. Similarly, since $\{Im(x_{n_k})\}_{k\in N_+}$ is a bounded sequence of $\mathbb{R}$, it has a convergent subsequence $\{Im(x_{n_{k_m}})\}_{m\in N_+}$. Then, the sequence $\{x_{n_{k_m}}\}_{m\in N_+}$ is a convergent subsequence of $\{x_n\}_{n\in N_+}$.
Q.E.D.

We have also shown in the course of proving the above lemma that any compact set in a metric space has a countably dense subset, and that it is covered by countably many open balls centered at some point in the countably dense subset. We refer to this property as the separability of compact sets in metric spaces, and it will repeatedly come in handy below.

The following is the characterization of relative compactness in metric spaces:

Theorem 1.16 Let (E,d) be a metric space, and A a subset of E. Then, A is relatively compact if and only if any sequence in A has a convergent subsequence.

Proof) Suppose that A is relatively compact. By definition, the closure $\overline{A}$ of A is compact; by the previous result, this means that $\overline{A}$ is sequentially compact. Any sequence in A is also a sequence in $\overline{A}$, so sequential compactness tells us that the sequence has a convergent subsequence.

Conversely, suppose that any sequence in A has a convergent subsequence. We will show that $\overline{A}$ is sequentially compact; by the previous result, this implies the compactness of $\overline{A}$ and thus the relative compactness of A.
Choose any sequence $\{x_n\}_{n\in N_+}$ in $\overline{A}$. For any $n \in N_+$, since $x_n \in \overline{A}$ there exists an $y_n \in A$ such that

$$d(y_n,x_n) < \frac{1}{n}.$$

Now consider the sequence $\{y_n\}_{n\in N_+}$ in A. By assumption, $\{y_n\}_{n\in N_+}$ has a convergent subsequence $\{y_{n_k}\}_{k\in N_+}$; letting $x \in E$ be the limit of the sequence, because $\{y_{n_k}\}_{k\in N_+}$ takes values in A, it must be the case that $x \in \overline{A}$.
It now holds that, for any $k \in N_+$,

$$d(x_{n_k},x) \leq d(x_{n_k},y_{n_k}) + d(y_{n_k},x) < \frac{1}{n_k} + d(y_{n_k},x).$$

Taking $k \to \infty$ on both sides now tells us that $\{x_{n_k}\}_{k\in N_+}$ is a subsequence of $\{x_n\}_{n\in N_+}$ that converges to $x \in \overline{A}$. Therefore, by definition, $\overline{A}$ is sequentially compact.
Q.E.D.

1.9 Continuous Functions

Let (E,τ) and (F,s) be topological spaces, and $f: E \to F$ a function mapping E into F. f is called a continuous function relative to τ and s if:

$$\text{For any } A \in s, \text{ the inverse image } f^{-1}(A) \in \tau.$$

Note that, if s is generated by some base $\mathbb{B}$ of F, then it suffices to check the above definition for the elements of $\mathbb{B}$. To see this, assume that $f^{-1}(B) \in \tau$ for every $B \in \mathbb{B}$. Then, for any $A \in s$, because A can be expressed as the aribtrary union $\bigcup_i B_i$ of sets in $\mathbb{B}$ by theorem 1.3, it follows that

$$f^{-1}(A) = \bigcup_i f^{-1}(B_i) \in \tau,$$

where the last inclusion follows because each $f^{-1}(B_i) \in \tau$ and topologies are closed under arbitrary unions. Therefore, f is continuous relative to τ and s by definition.
The following are some alternative characterizations of continuous functions.

Lemma 1.17 Let (E,τ) and (F,s) be topological spaces and $f: E \to F$ a function. Then, the following are equivalent:

i) f is continuous relative to τ and s.

ii) $f^{-1}(B)$ is closed in τ for any $B \subset F$ closed in s.

iii) For any $x \in E$ and neighborhood $V \in s$ of $f(x)$, there exists a neighborhood $U \in \tau$ of x such that $f(U) \subset V$.

Proof) Suppose f is continuous relative to τ and s. Then, for any subset B of F that is closed under s, since $B^c = F \setminus B \in s$, we can see that

$$f^{-1}(B) = f^{-1}(F \setminus (F \setminus B)) = f^{-1}(F) \setminus f^{-1}(B^c) = E \setminus f^{-1}(B^c) = f^{-1}(B^c)^c.$$

Since $B^c \in s$, $f^{-1}(B^c) \in \tau$ by continuity, and as such $f^{-1}(B^c)^c$ is closed under τ. This holds for any closed set B under s, so that *i)* implies *ii)*.
Conversly, suppose that $f^{-1}(B)$ is closed under τ for any $B \subset F$ closed under s. Then, for any $A \in s$, $A^c = F \setminus A$ is closed under s and

$$f^{-1}(A) = f^{-1}(F \setminus (F \setminus A)) = f^{-1}(F) \setminus f^{-1}(A^c) = E \setminus f^{-1}(A^c) = f^{-1}(A^c)^c,$$

where $f^{-1}(A^c)$ is closed under τ by hypothesis and thus $f^{-1}(A^c)^c \in \tau$. This holds for any $A \in s$, and as such, f is continuous relative to τ and s by definition.
Therefore, the statements *i)* and *ii)* are equivalent.

Now suppose that f is continuous relative to τ and s, and choose any $x \in E$ and a neighborhood $V \in s$ around $f(x)$. Then, because $f^{-1}(V) \in \tau$ and $x \in f^{-1}(V)$, letting $U = f^{-1}(V)$, $U \in \tau$ is a neighborhood around x. In addition, for any $y \in f(U)$, there exists a $z \in U$ such that $f(z) = y$; since $z \in f^{-1}(V)$ and thus $f(z) \in V$, it follows that $y \in V$ and thus $f(U) \subset V$. Therefore, *i)* implies *iii)*.
Conversely, suppose that, *iii)* holds. For any $A \in s$, if $f^{-1}(A) = \emptyset$, then $f^{-1}(A) \in \tau$ trivially. On the other hand, if $f^{-1}(A) \neq \emptyset$, then we can choose some $x \in f^{-1}(A)$; since $f(x) \in A$, A is a neighborhood around $f(x)$ and thus, by hypothesis, there exists a neighborhood $N_x \in \tau$ around x such that $f(N_x) \subset A$, or $N_x \subset f^{-1}(A)$. This holds for any $x \in f^{-1}(A)$, so

$$f^{-1}(A) = \bigcup_{x \in f^{-1}(A)} N_x \in \tau,$$

where the last inclusion follows because each $N_x \in \tau$ and topologies are closed under arbitrary unions. Therefore, for any $A \in s$, $f^{-1}(A) \in \tau$ and f is continuous relative to τ and s by definition.
Q.E.D.

We can also show that continuity is preserved under composition.

Lemma 1.18 Let (E,τ), (F,s) and (G,γ) be topological spaces, and $f : E \to F$, $g : F \to G$ functions that are continuous relative to τ and s, and s and γ, respectively. Then, the function $h : E \to G$ defined as $h = g \circ f$ is continuous relative to τ and γ.

Proof) Let $A \in \gamma$. By the definition of continuity, $g^{-1}(A) \in s$. Likewise, because f is continuous, $f^{-1}(g^{-1}(A)) \in \tau$. Therefore,

$$h^{-1}(A) = (g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A)) \in \tau.$$

This holds for any $A \in \gamma$, so by definition, h is continuous relative to τ and γ.
Q.E.D.

Many of the properties of continuous functions studied so far will have direct counterparts when it comes to measurable functions. In this sense, topological spaces and topology-preserving functions (continuous functions) are analogous to measurable spaces and σ-algebra-preserving functions (measurable functions).

A very important property of continuous functions is that they preserve openness aover inverse images and compactness over images. This is shown below:

Lemma 1.19 Let (E,τ) and (F,s) be topological spaces and $f:E\to F$ a continuous function. Then, for any compact set K in E, the image $f(K)\subset F$ is a compact set in F.

Proof) Let K be a compact set in E, and $\{V_\alpha\}\subset s$ an open cover of $f(K)$. It then holds that

$$K\subset\bigcup_\alpha f^{-1}(V_\alpha);$$

if $x\in K$, then $f(x)\subset f(K)\subset V_\alpha$ for some α, so $x\in f^{-1}(V_\alpha)$.
Because f is continuous, each of the sets $f^{-1}(V_\alpha)$ is an open set in τ. This means that $\{f^{-1}(V_\alpha)\}$ forms an open cover of K, and by the compactness of K, there exist $\alpha_1,\cdots,\alpha_n$ such that

$$K\subset f^{-1}(V_{\alpha_1})\cup\cdots\cup f^{-1}(V_{\alpha_n}).$$

Then,

$$f(K)\subset V_{\alpha_1})\cup\cdots\cup V_{\alpha_n};$$

for any $y\in f(K)$, there exists an $x\in K$ such that $y=f(x)$, and because $x\in f^{-1}(V_{\alpha_i})$ for some $1\le i\le n$, it follows that $y=f(x)\in V_{\alpha_i}$.
Thus, $\{V_\alpha\}$ has a finite subcover of $f(K)$, which implies that $f(K)$ is compact in F.
Q.E.D.

Consider metric spaces (E,d), (F,ρ) and let τ, s be the metric topologies on E, F induced by d, ρ. Then, we can show that the familiar $\varepsilon-\delta$ definition of continuity holds.

Lemma 1.20 Let $f: E \to F$ be a function. Then, the following are equivalent.

i) f is continuous relative to τ and s

ii) For any $x \in E$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\rho(f(x), f(y)) < \varepsilon \quad \text{for any } y \in E \text{ such that } d(x,y) < \delta.$$

Proof) Suppose f is continuous relative to τ and s. Choose any $x \in E$ and $\varepsilon > 0$; then, $B_\rho(f(x),\varepsilon) \in s$, so that

$$f^{-1}(B_\rho(f(x),\varepsilon)) \in \tau.$$

Because $f(x) \in B_\rho(f(x),\varepsilon)$, $x \in f^{-1}(B_\rho(f(x),\varepsilon))$, and by lemma 1.6, there exists a $\delta > 0$ such that $B_d(x,\delta) \subset f^{-1}(B_\rho(f(x),\varepsilon))$. This means that, for any $y \in E$ such that $d(x,y) < \delta$, or $y \in B_d(x,\delta)$, we have $y \in f^{-1}(B_\rho(f(x),\varepsilon))$, or $f(y) \in B_\rho(f(x),\varepsilon)$, that is, $\rho(f(x), f(y)) < \varepsilon$.
This shows that *i)* implies *ii)*.

Conversely, suppose that *ii)* holds. Choose any $A \in s$. If $f^{-1}(A) = \emptyset$, then $f^{-1}(A) \in \tau$ trivially.
Suppose that $f^{-1}(A) \neq \emptyset$. Then, there exists an $x \in f^{-1}(A)$, or $f(x) \in A$; by lemma 1.6, there exists an $\varepsilon > 0$ such that $B_\rho(f(x),\varepsilon) \subset A$. By hypothesis, there then also exists a $\delta > 0$ such that $B_d(x,\delta) \subset f^{-1}(B_\rho(f(x),\varepsilon))$. Denoting $N_x = B_d(x,\delta)$, it follows that

$$N_x =_d (x,\delta) \subset f^{-1}(B_\rho(f(x),\varepsilon)) \subset f^{-1}(A).$$

This holds for any $x \in f^{-1}(A)$, so

$$f^{-1}(A) = \bigcup_{x \in f^{-1}(A)} N_x \in \tau,$$

where the last inclusion follows because each $N_x \in \tau$ and topologies are closed under arbitrary unions. Therefore, for any $A \in s$, $f^{-1}(A) \in \tau$ and f is continuous relative to τ and s by definition.
Q,E.D.

1.10 Homeomorphisms

Let (E,τ) and (F,s) be topological spaces. A function $f:E\to F$ is said to be a homeomorphism if

i) f is continuous relative to τ and s

ii) f is a bijection, so that its inverse $f^{-1}:E\to F$ exists

iii) f^{-1} is continuous relative to s and τ.

If f is a homeomorphism, then, for any open $B\in s$, the inverse mapping $f^{-1}(B)\in\tau$ by the continuity of f. However, unlike ordinary continuous functions, the image of open sets under a homeomorphism is also open, since, for any $A\in\tau$,

$$f(A)=(f^{-1})^{-1}(A)\in s$$

by the continuity of f^{-1}.

Heuristically, the above means that any topological property of a set in E is preserved under a homeomorphism.
For instance, if $A\in\tau$, then $f(A)\in s$, while if $f(A)\in s$, then $f^{-1}(f(A))=A\in\tau$, so that A is open in E if and only if $f(A)$ is open in F.
Likewise, an arbitrary subset B of E is closed in E if and only if $f(B)$ is closed in E as well.
Furthermore, if K is compact in E, then $f(K)$ is compact in F by the property of continuous functions, while if $f(K)$ is compact in F, then $f^{-1}(f(K))=K$ is compact in E, again by continuity. This tells us that a subset K of E is compact if and only if $f(K)$ is compact in F.

The preservation of topological properties under homeomorphisms also extends to the continuity of functions. Consider topological spaces (E,τ), (F,s) and (G,γ), and suppose that there exists a homeomorphism $h:F\to G$. Then, if $f:E\to F$ is continuous relative to τ and s, the function $h\circ f:E\to G$ is continuous by the preservation of continuity across compositions, and if $h\circ f:E\to G$ is continuous relative to τ and γ, then

$$f=h^{-1}\circ(h\circ f)$$

is continuous relative to τ and s because h^{-1} is continuous relative to γ and s. We have thus seen that f is continuous if and only if $h\circ f$ is.

We have seen that a homeomorphism between two topological spaces (E,τ) and (F,s) allow them, in a sense, to share topological properties. As such, we might identify these spaces, or at least say that the spaces are homeomorphic.

1.10.1 The Standard Topology on $\mathbb{C}$

As an application of the results on homeomorphisms, we conside topologies on the complex plane. Letting $|\cdot|_c$ be the euclidean norm on $\mathbb{C}$, defined as

$$|z|_c = \sqrt{a^2+b^2}$$

for any $z = a+ib \in \mathbb{C}$, we define the standard topology $\tau_{\mathbb{C}}$ on $\mathbb{C}$ as the metric topology induced by the metric induced by the above norm.
The norm $|\cdot|_c$ is very reminiscent of the euclidean norm on $\mathbb{R}^2$, which assigns the value $\sqrt{a^2+b^2}$ to the 2-dimensional vector (a,b). Indeed, letting ρ and d^2 be the metrics induced by the euclidean norms on $\mathbb{C}$ and $\mathbb{R}^2$, and denoting the euclidean topology on $\mathbb{R}^2$ induced by d^2 as $\tau_{\mathbb{R}^2}$, we can see that the topological spaces $(\mathbb{C}, \tau_{\mathbb{C}})$ and $(\mathbb{R}^2, \tau_{\mathbb{R}^2})$ are homeomorphic.

To see this, let ρ and d^2 be the metrics induced by the euclidean norms on $\mathbb{C}$ and $\mathbb{R}^2$. We first define the function $f : \mathbb{C} \to \mathbb{R}^2$ as

$$f(z) = (Re(z), Im(z))$$

for any $z \in \mathbb{C}$. We can now show that f is a homeomorphism:

Lemma 1.21 The function f defined above is a homeomorphism of $(\mathbb{C}, \tau_{\mathbb{C}})$ and $(\mathbb{R}^2, \tau_{\mathbb{R}^2})$.

Proof) We first prove that f is continuous by using the $\varepsilon - \delta$ characterization of continuity on metric spaces.

For any $u, z \in \mathbb{C}$, note that

$$d^2(f(u), f(z)) = |(Re(u), Im(u)) - (Re(z), Im(z))| = |z-u|_c = \rho(u,z).$$

This means that, for any $\varepsilon > 0$, $\rho(u,z) < \varepsilon$ implies

$$d^2(f(u), f(z)) = \rho(u,z) < \varepsilon,$$

so f is uniformly continuous on $\mathbb{C}$. In fact, it is Lipschitz continuous with Lipschitz constant equal to 1.

Note also that f is a bijection. If $f(z) = f(u)$ for some $z, u \in \mathbb{C}$, then $Re(z) = Re(u)$ and $Im(z) = Im(u)$, so that $z = u$; f is an injective mapping. On the other hand, for any $(a,b) \in \mathbb{R}^2$, $f(a+ib) = (a,b)$, so f is also a surjective mapping onto $\mathbb{R}^2$, which shows that f is a bijection. Therefore, it has an inverse function $f^{-1} : \mathbb{R}^2 \to \mathbb{C}$.

We can now show that the inverse mapping f^{-1} is continuous relative to $\tau_{\mathbb{R}^2}$ and $\tau_{\mathbb{C}}$, again through the $\varepsilon-\delta$ characterization. As in the case of f, for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^2$,

$$\rho(f^{-1}(\mathbf{x}),f^{-1}(\mathbf{y}))=|(x_1+i\cdot x_2)-(y_1+i\cdot y_2)|_c=|\mathbf{x}-\mathbf{y}|=d^2(\mathbf{x},\mathbf{y}),$$

so f^{-1} is also Lipschitz continuous with Lipschitz constant equal to 1.
By definition, f is a homeomorphism.
Q.E.D.

Denote the above homeomorphism by $\phi:\mathbb{C}\to\mathbb{R}^2$. Due to the preservation of continuity under homeomorphisms, for any topological space (E,τ) and function $f:E\to\mathbb{C}$, f is continuous relative to τ and $\tau_{\mathbb{C}}$ if and only if $\phi\circ f:E\to\mathbb{R}^2$ is continuous relative to τ and $\tau_{\mathbb{R}^2}$. The function $\phi\circ f$ is specifically defined as

$$(\phi\circ f)(x)=\phi(f(x))=(Re(f(x)),Im(f(x)))$$

for any $x\in E$. This relationship allows us to obtain a very simple characterization of continuous complex functions:

Theorem 1.22 A function $f:E\to\mathbb{C}$ is continuous relative to τ and $\tau_{\mathbb{C}}$ if and only if $Re(f)$ and $Im(f)$ are continuous relative to τ and $\tau_{\mathbb{R}}$.

Proof) Denote $h=\phi\circ f$ for notational brevity.
Suppose that $Re(f)$ and $Im(f)$ are real-valued functions continuous relative to τ and $\tau_{\mathbb{R}}$. Then, for any $A,B\in\tau_{\mathbb{R}}$,

$$h^{-1}(A\times B)=(Re(f))^{-1}(A)\cap(Im(f))^{-1}(B)\in\tau$$

by the definition of continuity. Since the set of open rectangles is a base generating $\tau_{\mathbb{R}}^2=\tau_{\mathbb{R}^2}$, it follows that h is continuous relative to τ and $\tau_{\mathbb{R}^2}$, and therefore that f is continuous relative to τ and $\tau_{\mathbb{C}}$.

Conversely, suppose that f is continuous relative to τ and $\tau_{\mathbb{C}}$. Then, h is continuous relative to τ and $\tau_{\mathbb{R}^2}$. For any $A\in\tau_{\mathbb{R}}$,

$$(Re(f))^{-1}(A)=(Re(f))^{-1}(A)\cap(Im(f))^{-1}(\mathbb{R})=h^{-1}(A\times\mathbb{R})\in\tau,$$

where $(Im(f))^{-1}(\mathbb{R})=E$ and $A\times\mathbb{R}$ is an open rectangle on $\mathbb{R}^2$. It follows that $Re(f)$ is continuous relative to τ and $\tau_{\mathbb{R}}$, and by the same process, so is $Im(f)$.
Q.E.D.

1.11 Real and Complex Continuous Functions

1.11.1 Preservation of Continuity under Arithmetic Operations

Let (E,τ) be a topological space, and let $f : E \to \mathbb{R}$ be a numerical function. By the previous observaiton that a necessary and sufficient condition for continuity is for the inverse image of any member of a base generating $\tau_{\mathbb{R}}$ to be contained in τ, we have that f is continuous if and only if $f^{-1}((a,b)) \in \tau$ for any open interval (a,b) with rational endpoints, since the collection of such sets generates $\tau_{\mathbb{R}}$.

Real-valued continuous functions are of special interest in that the addition, scalar multiplication and product of such functions are still continuous real valued functions. This is stated below:

Lemma 1.23 Let (E,τ) be a topological space, and let $f,g : E \to \mathbb{R}$ be real-valued functions continuous relative to τ and the standard topology $\tau_{\mathbb{R}}$ on $\mathbb{R}$. The following hold true:

i) $f+g$ is a continuous function.

ii) For any $c \in \mathbb{R}$, cf is a continuous function.

iii) The product fg is a continuous function.

Proof) Define $h : \mathbb{R}^2 \to \mathbb{R}$ as

$$h(x,y) = x+y$$

for any $(x,y) \in \mathbb{R}^2$. We will show that h is continuous relative to $\tau_{\mathbb{R}}^2$ and $\mathbb{R}$.

Choose any open interval $(a,b) \subset \mathbb{R}$ with rational endpoints. Then,

$$\begin{aligned} h^{-1}((a,b)) &= \{(x,y) \in \mathbb{R}^2 \mid a < h(x) < b\} = \{(x,y) \in \mathbb{R}^2 \mid a < x+y < b\} \\ &= \{(x,y) \in \mathbb{R}^2 \mid a-y < x < b-y\} \\ &= \{(x,y) \in \mathbb{R}^2 \mid a-y < x\} \cap \{(x,y) \in \mathbb{R}^2 \mid x < b-y\} \\ &= \bigcup_{r \in \mathbb{R}} \left([(r,+\infty) \times \ (a-r,+\infty)] \cap [(-\infty,b-r) \times (-\infty,r)]\right). \end{aligned}$$

Each set on the right hand side is the intersection of two open rectangles, so each such set is in $\tau_{\mathbb{R}}^2$, and because topologies are closed under unions, $h^{-1}((a,b)) \in \tau_{\mathbb{R}}^2$.
Because the set of all open intervals with rational endpoints is a base of $\mathbb{R}$ generating $\tau_{\mathbb{R}}$, by the characterization of continuity stated above, h is continuous relative to $\tau_{\mathbb{R}}^2$ and $\mathbb{R}$.

Now define the function $\phi : E \to \mathbb{R}^2$ as

$$\phi(x) = (f(x), g(x))$$

for any $x \in E$. Then, for any $A, B \in \tau_{\mathbb{R}}$,

$$\phi^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B) \in \tau$$

by the continuity of f and g, and because the set of all open rectangles is a base on $\mathbb{R}^2$ generating $\tau_{\mathbb{R}}^2$, it follows that ϕ is continuous relative to τ and $\tau_{\mathbb{R}}^2$.

Finally,

$$f+g = h(f,g) = h \circ \phi;$$

by the preservation of continuity across compositions, $f+g$ is continuous relative to τ and $\tau_{\mathbb{R}}$.

Now choose some $c \in \mathbb{R}$. If $c = 0$, then $cf = 0$ is trivially continuous. Assume that $c > 0$. Then, for any open interval $(a,b) \subset \mathbb{R}$,

$$(cf)^{-1}((a,b)) = \{x \in E \mid a < cf(x) < b\} = \{x \in E \mid \frac{a}{c} < f(x) < \frac{b}{c}\} = f^{-1}\left(\left(\frac{a}{c}, \frac{b}{c}\right)\right) \in \tau,$$

so that cf is continuous relative to τ and $\tau_{\mathbb{R}}$.
If $c < 0$, then similarly, for any open interval $(a,b) \subset \mathbb{R}$,

$$(cf)^{-1}((a,b)) = \{x \in E \mid a < cf(x) < b\} = \{x \in E \mid \frac{b}{c} < f(x) < \frac{a}{c}\} = f^{-1}\left(\left(\frac{b}{c}, \frac{a}{c}\right)\right) \in \tau,$$

so that cf is again continuous relative to τ and $\tau_{\mathbb{R}}$.

Finally, note that

$$fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right).$$

Since the mapping $x \mapsto x^2$ is continuous on $\mathbb{R}$, and continuity is preserved across compositions and over addition and scalar multiplication, we can see that fg is continuous relative to τ and $\tau_{\mathbb{R}}$.
Q.E.D.

Let $f : E \to \mathbb{C}$ be a complex-valued function. We saw in the previouss section that f is continuous relative to τ and $\tau_{\mathbb{C}}$ (denoted in brief as "continuous relative to τ") if and only if its real and imaginary parts $Re(f), Im(f)$ are continuous relative to τ and $\tau_{\mathbb{R}}$. The following shows that the sum, scalar multiple and product of continuous complex functions are also continuous:

Lemma 1.24 Let (E, τ) be a topological space, and let $f, g : E \to \mathbb{C}$ be complex-valued functions continuous relative to τ. The following hold true:

i) For any $c \in \mathbb{C}$, $cf + g$ is a continuous function.

ii) The product fg is a continuous function.

Proof) For any $c \in \mathbb{C}$, letting $c = a + ib$ for $a, b \in \mathbb{R}$, we have

$$\begin{aligned} cf + g &= (a + ib) \cdot (Re(f) + i \cdot Im(f)) + Re(g) + i \cdot Im(g) \\ &= (aRe(f) - bIm(f)) + i \cdot (bRe(f) + aIm(f)) + Re(g) + i \cdot Im(g) \\ &= (aRe(f) - bIm(f) + Re(g)) + i \cdot (bRe(f) + aIm(f) + Im(g)). \end{aligned}$$

Since the continuity of real valued functions is preserved over addition and scalar multiplication, $aRe(f) - bIm(f) + Re(g)$ and $bRe(f) + aIm(f) + Im(g)$ are continuous relative to τ and $\tau_{\mathbb{R}}$. It follows that $cf + g$ is a complex function continuous relative to τ.

As for the product, note that

$$\begin{aligned} fg &= (Re(f) + i \cdot Im(f))(Re(g) + i \cdot Im(g)) \\ &= Re(f)Re(g) - Im(f)Im(g) + i \cdot (Im(f)Re(g) + Re(f)Im(g)). \end{aligned}$$

Since the continuity of real valued functions is preserved over products, $Re(f)Re(g) - Im(f)Im(g)$ and $Im(f)Re(g) + Re(f)Im(g)$ are continuous relative to τ and $\tau_{\mathbb{R}}$. It follows that fg is a complex function continuous relative to τ.
Q.E.D.

1.11.2 Semicontinuity

We say that a real-valued function $f : E \to \mathbb{R}$ is upper semicontinuous if

$$\{x \in E \mid f(x) < a\} = f^{-1}((-\infty, a)) \in \tau$$

for any $a \in \mathbb{R}$.
Likewise, we say that f is lower semicontinuous if

$$\{x \in E \mid f(x) > a\} = f^{-1}((a, +\infty)) \in \tau$$

for any $a \in \mathbb{R}$.
Note that f is continuous if and only if it is both upper and lower semicontinuous: if f is continuous, then because $(a, +\infty)$ and $(-\infty, a)$ are all open sets in $\tau_{\mathbb{R}}$ for any $a \in \mathbb{R}$,

$$f^{-1}((-\infty, a)), f^{-1}((a, +\infty)) \in \tau$$

for any $a \in \mathbb{R}$, so that f is both upper and lower semicontinuous, while if f is both upper and lower semicontinuous, then for any rational $a, b \in \mathbb{Q}$ such that $a < b$,

$$f^{-1}((a, b)) = f^{-1}((-\infty, b)) \cap f^{-1}((a, +\infty)) \in \tau,$$

so that f is continuous.

The following result shows that the indicator of open sets and closed sets are lower and upper semicontinuous, respectively:

Lemma 1.25 Let (E, τ) be a topological space. The following hold true:

i) For any open set $A \in \tau$, the indicator I_A is lower semicontinuous.

ii) For any closed subset A of E, the indicator I_A is upper semicontinuous.

Proof) i) Suppose $A \in \tau$. Then, for any $a \in \mathbb{R}$,

$$(I_A)^{-1}((a, +\infty)) = \begin{cases} \emptyset & \text{if } 1 \leq a \\ A & \text{if } 0 \leq a < 1 \\ E & \text{if } a < 0 \end{cases},$$

so that $(I_A)^{-1}((a, +\infty)) \in \tau$ and I_A is lower semicontinuous.

ii) Suppose A is closed, so that $A^c \in \tau$. Then, for any $a \in \mathbb{R}$,

$$(I_A)^{-1}((-\infty, a)) = \begin{cases} E & \text{if } 1 < a \\ A^c & \text{if } 0 < a \leq 1 \\ \emptyset & \text{if } a \leq 0 \end{cases},$$

so that $(I_A)^{-1}((a, +\infty)) \in \tau$ and I_A is upper semicontinuous.

Q.E.D.

We can also see that upper and lower semicontinuity are preserved under suprema and infima:

Lemma 1.26 Let (E, τ) be a topological space, and $\{f_n\}_{n\in N_+}$ a sequence of real functions on E such that $\sup_{n\in N_+} f_n(x)$ and $\inf_{n\in N_+} f_n(x)$ exist in $\mathbb{R}$ for any $x \in E$. The following hold true:

i) If each f_n is lower semicontinuous, $\sup_{n\in N_+} f_n$ is also lower semicontinuous.

ii) If each f_n is upper semicontinuous, $\inf_{n\in N_+} f_n$ is also upper semicontinuous.

Proof) Suppose that $\{f_n\}_{n\in N_+}$ is a sequence of lower semicontinuous functions, and define $f = \sup_{n\in N_+} f_n$; note that f is real-valued by hypothesis. Then, for any $a \in \mathbb{R}$,

$$f^{-1}((a, +\infty)) = \{x \in E \mid f(x) > a\} = \bigcup_n \{x \in E \mid f_n(x) > a\},$$

and because each $\{x \in E \mid f_n(x) > a\} \in \tau$ by lower semicontinuity, it follows that $f^{-1}((a, +\infty)) \in \tau$ as well. Therefore, f is lower semicontinuous.

On the other hand, suppose $\{f_n\}_{n\in N_+}$ is a sequence of upper semicontinuous functions, and define $f = \inf_{n\in N_+} f_n$; note that f is real-valued by hypothesis. Then, for any $a \in \mathbb{R}$,

$$f^{-1}((-\infty, a)) = \{x \in E \mid f(x) < a\} = \bigcup_n \{x \in E \mid f_n(x) < a\},$$

and because each $\{x \in E \mid f_n(x) < a\} \in \tau$ by upper semicontinuity, it follows that $f^{-1}((-\infty, a)) \in \tau$ as well. Therefore, f is upper semicontinuous.
Q.E.D.

1.12 Connectedness

Another important topological property, of great importance even outside the context of metric spaces, is connectedness. Let (E,τ) be a topological space. Any two subsets A,B of E are said to be separated if $\overline{A}\cap B=\emptyset$ and $A\cap\overline{B}=\emptyset$, that is, if each set does not intersect with the closure of the other. A subset of E is said to be connected if it cannot be represented as the union of two nonempty separated sets. The empty set is trivially connected.

A related notion is path-connectedness. A path between two points x,y in a topological space (E,τ) is defined as a continuous function (with respect to the standard topology on $\mathbb{R}$ and τ) $f:[0,1]\to E$ such that $f(0)=x$ and $f(1)=y$. A non-empty subset A of E is said to be path-connected if there exists a path between any two points x,y in A.

Also important to the study of connected sets is the concept of a linear continuum, which generalizes the properties of intervals and rays on the real line that render it connected. An ordered set L is said to be a linear continuum if

1) It has the least upper bound property, and

2) For any $x,y\in L$ such that $x<y$, there exists a $z\in L$ such that $x<z<y$.

Note that any interval or ray on the real line is a linear continuum. There exists a very convenient characterization of connected subsets of a linear continuum.

Theorem 1.27 Let L be a linear continuum and τ the order topology on L. A nonempty subset K of L is connected if and only if, for any $x,y\in K$ such that $x<y$, $x<z<y$ implies $z\in K$.

Proof) **Necessity**

Let K be a subset of L, and suppose that there exist $x,y\in K$ and $z\in L$ such that $x<z<y$ but $z\notin K$. Define

$$A=K\cap(-\infty,z),\quad B=K\cap(z,+\infty),$$

where $(-\infty,z)$ is shorthand for the open ray $\{w\in L\mid w<z\}$. Then, $K=A\cup B$, and A,B are non-empty sets (they contain x and y, respectively). Furthermore, suppose that $w\in\overline{A}\cap B$. Note that $w\in B$, so that $z<w$. In addition,

$$w\in\overline{A}\subset(-\infty,z]:=\{u\in L\mid u\le z\}.$$

We are left with $w\le z<w$, a contradiction, so it must be the case that $\overline{A}\cap B=\emptyset$. A symmetric process reveals that $A\cap\overline{B}=\emptyset$, and it follows that K is the union of nonempty separated sets A and B. Thus, K is not connected.

Sufficiency

Let K be a subset of L, and suppose that K is not connected. Then there exists nonempty separated sets $A, B \subset L$ such that $K = A \cup B$. Choose $x \in A$ and $y \in B$, and assume without loss of generality that $x < y$ (they cannot be equal since A and B are disjoint). Define the set $C = A \cap [x, y]$; C is a nonempty (contains x) and bounded subset of L (bounded above by y), so by the least upper bound property of L, C has a supremum α in L.

By definition, $x \leq \alpha \leq y$, and examine what happens when $\alpha \in B$. Since the supremum of a set is contained in its closure, we can see that $\alpha \in \overline{C} \subset \overline{A}$. Under the above supposition, this means that α is in the intersection of $\overline{A}$ and B, which contradicts the fact that A and B are separated. Therefore, $\alpha \notin B$ and $\alpha < y$.

We now consider two cases. If $\alpha \notin A$, then $x < \alpha < y$ and α is not contained in either A or B, meaning that it is not contained in K. Thus, in this case, we put $z = \alpha$. On the other hand, if $\alpha \in A$, then $\alpha \notin \overline{B}$ by the separation property. This means that α is not a limit point of B, that is, there exists some neighborhood V of α such that $V \cap B = \emptyset$. The collection of all rays and open intervals in L form a base generating the order topology τ of L, so there must exist a ray or open interval $b \subset L$ such that $\alpha \in b \subset V$. We consider the following cases:

- **Case 1**: $b = \{w \in L \mid r < w\}$ for some $r \in L$
  In this case, b is the ray bounded below by r. Since $\alpha \in b$, it must be the case that $b < \alpha$, and because $\alpha < y$, y is also contained in the ray b. By implication, $y \in V$ and $y \notin B$, a contradiction. Therefore, b cannot be a ray that is bounded below.

- **Case 2**: $b = \{w \in L \mid w < r\}$ for some $r \in L$
  In this case, b is the ray bounded above by r. This indicates that $\alpha < r$, and if $y < r$, we must have $y \in V$ and $y \notin B$, another contradiction. Therefore, $r \leq r$, and by the property of linear continuums, there exists a $z \in L$ such that $\alpha < z < y \leq r$; this z is contained in b, so it is not an element of B.

- **Case 3**: $b = (r, s)$ for some $r, s \in L$
  In this case, b is an open interval, and $r < \alpha < s$. Again, $y < s$ implies that $y \notin B$, so we must have $r < \alpha < s \leq y$. By the same process as above, this allows us to choose a $z \in L$ such that $\alpha < z < y$ and $z \notin B$.

We have shown that, in any case, we can choose some $z \in L$ such that $x < \alpha < z < y$ and $z \notin B$. This z cannot be an element of A, since in this case it is contained in C and we have the contradiction $\alpha = \sup C < z$. As such, we have found a $z \in L$ such that $x < z < y$ and $z \notin K$.

Q.E.D.

The above result shows us that the real line equipped with its standard topology, as well as any intervals or rays contained in $\mathbb{R}$, are all connected. Connectedness is also a property, along with compactness, that is preserved by the image of continuous functions, as we show below:

Theorem 1.28 Let (E,τ) and (F,s) be topological spaces and $f: E \to F$ a continuous function. If E is a connected space, then the image $f(E) \subset F$ is a connected subset of F.

Proof) Suppose that the image $f(E)$ is not connected. Then, there exist nonempty separated sets $A, B \subset F$ such that $f(E) = A \cup B$. Define $A_0 = f^{-1}(A)$ and $B_0 = f^{-1}(B)$. By the continuity of f, the inverse image $f^{-1}(\overline{A})$ is a closed set containing $f^{-1}(A)$, so it follows that $\overline{A_0} \subset f^{-1}(\overline{A})$. Therefore,

$$\overline{A_0} \cap B_0 \subset f^{-1}(\overline{A} \cap B) = \emptyset,$$

where the final equality follows because $\overline{A} \cap B = \emptyset$ by separatedness. Similarly, we can show that $A_0 \cap \overline{B_0} = \emptyset$, and A_0, B_0 are nonempty separated subsets of E. It follows that

$$E \subset f^{-1}(f(E)) = f^{-1}(A \cup B) = A_0 \cup B_0.$$

Conversely, because E is the domain of f, A_0 and B_0 are subsets of E. We can see that $E = A_0 \cup B_0$, or that it is the union of two nonempty separated subsets of E. By definition, E is not connected, and the claim of the theorem follows by contraposition.

Q.E.D.

The famous intermediate value theorem follows from combining the two preceding results.

Theorem 1.29 **(Intermediate Value Theorem)**
Let $f: [a,b] \to \mathbb{R}$ be a continuous function such that $f(a) < f(b)$. Then, for any $z \in \mathbb{R}$ such that $f(a) < z < f(b)$, there exists a $c \in (a,b)$ such that $f(c) = z$.

Proof) Since f is a continuous function and the closed interval $[a,b]$ is a connected set in the standard topology on the real line, the image $f([a,b])$ is a connected set, also in the standard topology. The points $f(a)$ and $f(b)$ are clearly contained in $f([a,b])$, and since $f([a,b])$ is a connected set on the real line, which is a linear continuum, by the characterization of connected sets on linear continuums proven above any $z \in \mathbb{R}$ such that $f(a) < z < f(b)$ is also contained in the connected set $f([a,b])$. Therefore, there must exist a $c \in (a,b)$ such that $f(c) = z$.

Q.E.D.

As a brief aside, we can actually prove the intermediate value theorem without taking recourse to connectedness by making use of the binary search algorithm. Intuitively, given a continuous function $f : [a,b] \to \mathbb{R}$ such that $f(a) < f(b)$ and some $f(a) < c < f(b)$, to find a $z \in (a,b)$ such that $f(z) = c$ we continuously bisect the interval (a,b) and move into the interval on which the sign of $f - c$ is different on the endpoints. Taking the limit then leads us to the desired value z.

Theorem 1.29 (Alternate Proof) (Intermediate Value Theorem)
Let $f : [a,b] \to \mathbb{R}$ be a continuous function such that $f(a) < f(b)$. Then, for any $z \in \mathbb{R}$ such that $f(a) < z < f(b)$, there exists a $c \in (a,b)$ such that $f(c) = z$.

Proof) We will construct two sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ taking values in $[a,b]$ as follows. First, we put $x_0 = a$ and $y_0 = b$. Suppose that we have defined $x_n \leq y_n$ for some $n \in \mathbb{N}$ so that $f(x_n) \leq c$ and $f(y_n) > c$. Defining $m_n = \frac{x_n+y_n}{2}$, the midpoint of the interval $[x_n, y_n]$ at which $f(x_n) \leq c$ and $f(y_n) \leq c$, we let

$$x_{n+1} = \begin{cases} m_n & \text{if } f(m_n) \leq c \\ x_n & \text{if } f(m_n) > c \end{cases}, \quad y_{n+1} = \begin{cases} y_n & \text{if } f(m_n) \leq c \\ m_n & \text{if } f(m_n) > c \end{cases}.$$

Heuristically, when we move from $[x_n, y_n]$ to $[x_{n+1}, y_{n+1}]$, we are moving onto the interval on which the first endpoint is below c and the other is above c. To see this formally, note that, if $f(m_n) \leq c$, then $f(x_{n+1}) = f(m_n) \leq c$, while $f(y_{n+1}) = f(y_n) > c$. On the other hand, if $f(m_n) > c$, then $f(x_{n+1}) = f(x_n) \leq c$, while $f(y_{n+1}) = f(m_n) > c$.

The sequence $\{[x_n, y_n]\}_{n\in\mathbb{N}}$ of intervals constructed inductively as above satisfies the condition

$$f(x_n) \leq c < f(y_n)$$

for any $n \in \mathbb{N}_+$. Since $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ take values in the compact set $[a,b]$, there exist subsequences $\{x_{n_k}\}_{k\in N_+}$ and $\{y_{n_k}\}_{k\in N_+}$ that converge to points x and y in $[a,b]$[1]. By the continuity of f,

$$f(x) = \lim_{n\to\infty} f(x_n) \leq c \quad \text{and} \quad f(y) = \lim_{n\to\infty} f(y_n) \geq c.$$

Furthermore, since for any $n \in N_+$ we have

$$|x_{n+1} - y_{n+1}| = \left|\frac{x_n - y_n}{2}\right|,$$

it follows that

$$|x_n - y_n| = \frac{1}{2}|x_{n-1} - y_{n-1}| = \cdots 2^{-n}|x_0 - y_0| = 2^{-n}(b-a)$$

[1] The indices $\{n_k\}_{k\in N_+}$ can be taken as the same across the wo subsequences by first taking the subsequence of $\{x_n\}_{n\in\mathbb{N}}$ and then the subsequence of the subsequence of $\{y_n\}_{n\in\mathbb{N}}$ defined on the indices of the earlier subsequence.

for any $n \in N_+$. Taking $n \to \infty$ on both sides, we can see that

$$\lim_{n\to\infty} |x_n - y_n| = 0,$$

so that their limits x and y are equal; put $x = y = z \in [a,b]$. It follows that

$$c \le f(z) \le c,$$

so that $f(z) = c$. Finally, since $f(a) < f(z) = c < f(b)$, it must be the case that $z \in (a,b)$.

Q.E.D.

We conclude our study of connectedness by stating the relationship between path-connectedness and connectedness. The notion of path-connectedness also gives rise to a generalization of the intermediate value theorem.

Theorem 1.30 The following hold true:

i) Path-connected sets are connected.

ii) Convex sets are path-connected.

iii) **(Generalization of Intermediate Value Theorem)**
Let (E,τ) be a topological space and A a path connected subset of A. Let $f : A \to \mathbb{R}$ be a continuous function, and $a,b \in A$ points such that $f(a) < f(b)$. Then, for any $c \in \mathbb{R}$ such that $f(a) < c < f(b)$, there exists a $z \in A$ such that $f(z) = c$.

Proof) i) Let (E,τ) be a topological space, and let K be a (nonempty) path-connected subset of E. Suppose that K is not connected, so that there exist nonempty separated subsets A,B of E such that $K = A \cup B$. Choose $x \in A$ and $y \in B$. By the path-connectedness of K, there exists a continuous function $f : [0,1] \to K$ such that $f(0) = x$ and $f(1) = y$. Define

$$A_0 = f^{-1}(A), \quad B_0 = f^{-1}(B).$$

by the same process as the proof above, we can see that A_0 and B_0 are nonempty separated subsets of $[0,1]$ such that $[0,1] = A_0 \cup B_0$. However, this contradicts the connectedness of the interval $[0,1]$, so it must be the case that K is connected.

ii) Let $\mathbb{R}^n$ be the n-dimensional euclidean space with the euclidean topology, and K a convex subset of $\mathbb{R}^n$. Choose any $x,y \in K$ and define the function $f : [0,1] \to \mathbb{R}^n$

as

$$f(t) = tx + (1-t)y$$

for any $t \in [0,1]$. Then, f takes values in K by convexity and is trivially continuous. This holds for any $x, y \in K$, so K is path connected. In fact, by the above result, K is also connected.

iii) Let (E, τ) be a topological space and A a path connected subset of E. Let $f : A \to \mathbb{R}$ and $a, b \in A$ be chosen so that f is a continuous function such that $f(a) < f(b)$. Choose any $c \in \mathbb{R}$ such that $f(a) < c < f(b)$. We want to find a $z \in A$ such that $f(z) = c$.
By the path-connectedness of A, there exists a continuous function $g : [0,1] \to A$ such that $g(0) = a$ and $g(1) = b$. Defining $h = f \circ g : [0,1] \to \mathbb{R}$, h is continuous because continuity is preserved across compositions. Since

$$h(0) = f(g(0)) = f(a) < c < f(b) = f(g(1)) = h(1),$$

by the IVT there exists a $t \in (0,1)$ such that $h(t) = c$. Defining $z = g(t) \in A$, this immediately tells us that $f(z) = c$.

Q.E.D.

Putting together the last two claims of the above theorem, we can formulate a version of the intermediate value theorem for convex sets on euclidean spaces. Given a convex set $A \subset \mathbb{R}^n$ and a function $f : A \to \mathbb{R}$, suppose that $f(x) < f(y)$ for some $x, y \in A$. For any $c \in \mathbb{R}$ such that $f(x) < c < f(y)$, there then exists some $z \in A$ such that $f(z) = c$, since the convexity of A implies that it is path-connected.

1.13 Urysohn's Lemma for Locally Compact Hausdorff Spaces

Let (E,τ) be a topological space, and $f:E\to\mathbb{C}$ a complex continuous function relative to τ. The support of f is defined as the closure of the set $\{x\in E \mid f(x)\neq 0\}$, that is, the set on which it is non-zero. Now define the following set:

$$C_c(E,\tau)=\{f:E\to\mathbb{C}\mid f \text{ is continuous relative to } \tau \text{ and has compact support}\}$$

It is easy to see that $C_c(E,\tau)$ is a vector space over $\mathbb{C}$; because the space of all complex functions on E is known to be a vector space over $\mathbb{C}$, we need only verify that $C_c(E,\tau)$ is a subspace of the larger function space. To this end, note that the zero function is contained in $C_c(E,\tau)$ because it is trivially continuous and has support $\emptyset$, which is compact.
Furthermore, for any $a\in\mathbb{C}$ and $f,g\in C_c(E,\tau)$, $af+g$ is a complex continuous function relative to τ. To see that it has compact support, note that

$$\{f=0\}\cap\{g=0\}\subset\{cf+g=0\},$$

so that

$$\{cf+g\neq 0\}\subset\{f\neq 0\}\cup\{g\neq 0\};$$

because f,g have compact support, and the finite union of compact sets is compact, $\overline{\{f\neq 0\}}\cup\overline{\{g\neq 0\}}$ is a compact set in E. Thus, the support of $cf+g$ is a closed set contained in the compact set $\overline{\{f\neq 0\}}\cup\overline{\{g\neq 0\}}$; it follows that it is itself compact.
This shows that $cf+g$ is a continuous function with compact support and thus contained in $C_c(E,\tau)$. This in turn completes the proof that $C_c(E,\tau)$ is a subspace of the entire function space and thus a vector space over the complex field.

For any $f\in C_c(E,\tau)$, the fact that f has compact support imparts on f many useful properties. For instance, letting $A=\overline{\{f\neq 0\}}$ be the support of f, the range of f is given as $f(E)=f(A)\cup\{0\}$. Because the image of a compact set is compact, $f(A)$ is compact in $\mathbb{C}$, and since the singleton $\{0\}$ is also compact, $f(E)$ is compact.

Before proving the main result of this section, we define some notations:
For any set $K \subset E$ and function $f : E \to \mathbb{C}$, we denote $K \prec f$ to represent:

$$f \in C_c(E,\tau), \quad K \text{ is compact}, \quad 0 \leq f(x) \leq 1 \text{ for any } x \in E, \quad f(x) = 1 \text{ for any } x \in K.$$

Likewise, for any set $V \subset E$, we denote $f \prec V$ to represent:

$$f \in C_c(E,\tau), \quad V \text{ is open}, \quad 0 \leq f(x) \leq 1 \text{ for any } x \in E, \quad \text{the support of } f \text{ lies in } V.$$

Note that $\overline{\{f \neq 0\}} \subset V$ implies that $\{f \neq 0\} \subset V$ and thus that $f(x) = 0$ for any $x \notin V$.
We denote $K \prec f \prec V$ if $K \prec f$ and $f \prec V$.
The following is Urysohn's lemma for locally compact Hausdorff spaces, which shows that any indicator function can be approximated by continuous functions in $C_c(E,\tau)$ with an arbitrary degree of precision:

Theorem 1.31 (Urysohn's Lemma for Locally Compact Hausdorff Spaces)
Let (E,τ) be a locally compact Hausdorff space. For any compact set K and open set V such that $K \subset V$, there exists a function $f \in C_c(E,\tau)$ such that

$$K \prec f \prec V,$$

that is, f is a complex continuous function with compact support such that

$$f(x) \in \begin{cases} \{0\} & \text{if } x \notin V \\ [0,1] & \text{if } x \in V \setminus K \\ \{1\} & \text{if } x \in K \end{cases}.$$

By implication, $I_K \leq f \leq I_V$.

Proof) First, we arrange the elements of the countable set $\mathbb{Q} \cap (0,1)$ into the sequence $\{r_n\}_{n \in N_+}$, and put $r_{-1} = 1$, $r_0 = 0$. By theorem 1.14, because K is a compact set contained in the open set V, there exists a $V_1 \in \tau$ with compact closure such that $K \subset V_1 \subset \bar{V_1} \subset V$. Since $\bar{V_1}$ is itself a compact set contained in V, we can apply theorem 1.14 once again to find a $V_0 \in \tau$ such that $\bar{V_1} \subset V_0 \subset \bar{V_0} \subset V$. Putting these together, we have the result

$$K \subset V_1 \subset \bar{V_1} \subset V_0 \subset \bar{V_0} \subset V.$$

Suppose, for some $n \geq 0$, that we have found $V_{r_{-1}}, V_{r_0}, \cdots, V_{r_n} \in \tau$ with compact closure such that

$$\overline{V_{r_j}} \subset V_{r_i} \text{ if } r_j > r_i \text{ for any } -1 \leq i,j \leq n.$$

Since $r_{n+1} \in (0,1)$, there exist $-1 \leq i,j \leq n$ such that $r_i < r_{n+1} < r_j$ when $r_{-1}, \cdots, r_{n+1}$

are ordered in ascending order; the existence of such r_i and r_j can be ascertained by noting that r_{n+1} cannot be the smallest or largest member of $\{r_{-1}, \cdots, r_{n+1}\}$, since 0 and 1 are contained in it as well.
Since $\overline{V_{r_j}} \subset V_{r_i}$ by the inductive hypothesis ($r_i < r_j$), by theorem 1.14 there exists a $V_{r_{n+1}} \in \tau$ with compact closure such that

$$\overline{V_{r_j}} \subset V_{r_{n+1}} \subset \overline{V_{r_{n+1}}} \subset V_{r_i}.$$

Constructing the sequence $\{V_r\}_{r\in\mathbb{Q}\cap[0,1]}$ in this manner, we end up with a sequence of open sets such that

$$\overline{V_r} \subset V_q \text{ if } r > q \text{ for any } r, q \in \mathbb{Q}\cap[0,1] \text{ and}$$

$$K \subset V_1 \subset \bar{V_1} \subset V_0 \subset \bar{V_0} \subset V.$$

Using the sets constructed above, define the sequence $\{f_r\}_{r\in\mathbb{Q}\cap[0,1]}$ and $\{g_r\}_{r\in\mathbb{Q}\cap[0,1]}$ of functions on E as follows:

$$f_r(x) = \begin{cases} r & \text{if } x \in V_r \\ 0 & \text{if } x \notin V_r \end{cases} \quad \text{and} \quad g_r(x) = \begin{cases} 1 & \text{if } x \in \bar{V_r} \\ r & \text{if } x \notin \bar{V_r} \end{cases}$$

for any $x \in E$ and $r \in \mathbb{Q}\cap[0,1]$. Note that, for any $r \in \mathbb{Q}\cap[0,1]$, we can write

$$f_r = r \cdot I_{V_r} \quad \text{and} \quad g_r = I_{\bar{V_r}} + r \cdot I_{\bar{V_r}^c}.$$

For any $a \in \mathbb{R}$,

$$f_r^{-1}((a,+\infty)) = \begin{cases} \emptyset & \text{if } r \le a \\ V_r & \text{if } 0 \le a < r \\ E & \text{if } a < 0 \end{cases} \quad \text{and} \quad g_r^{-1}((-\infty,a)) = \begin{cases} E & \text{if } 1 < a \\ \bar{V}_r^c & \text{if } r < a \le 1 \\ \emptyset & \text{if } a \le r \end{cases},$$

so $f_r^{-1}((a,+\infty)), g_r^{-1}((-\infty,a)) \in \tau$. This implies that f_r is lower semicontinuous, while g_r is upper semicontinuous.
Define

$$f = \sup_{r\in\mathbb{Q}\cap[0,1]} f_r \quad \text{and} \quad g = \inf_{r\in\mathbb{Q}\cap[0,1]} g_r,$$

where f, g are well-defined and take values in $[0,1]$ because for each $r \in \mathbb{Q}\cap[0,1]$, f_r and g_r take values in $[0,1]$. Since f and g are the supremum and infimum of lower semicontinuous and upper semicontinuous real valued functions, respectively, by lemma 1.24 we have f and g are lower and upper semicontinuous functions.

We now study the properties of the function f: by design,

$$0 \leq f(x) \leq 1 \text{ for any } x \in E.$$

Moreover, for any $r \in \mathbb{Q} \cap [0,1]$, since $f_r(x) = 0$ for any $x \in V_r^c$, and $V_0^c \subset V_r^c$, it follows that

$$f(x) = \sup_{r \in \mathbb{Q} \cap [0,1]} f_r(x) = 0 \text{ for any } x \notin V_0.$$

This implies that $\{f \neq 0\} \subset V_0$, and because $\overline{\{f \neq 0\}} \subset \bar{V_0} \subset V$, where $\bar{V_0}$ is a compact set and $\overline{\{f \neq 0\}}$ a closed one, $\overline{\{f \neq 0\}}$ is compact. Thus, f has compact support that lies in the open set V.
Finally, if $x \in K$, then $x \in V_1 \subset V_r$ for any $r \in \mathbb{Q} \cap [0,1]$, which tells us that

$$f(x) = \sup_{r \in \mathbb{Q} \cap [0,1]} f_r(x) = \sup_{r \in \mathbb{Q} \cap [0,1]} r = 1.$$

It remains to see that f is continuous relative to τ; this can be done by showing that $f = g$, at which point f will be both upper and lower semicontinuous and thus continuous.
For any $r, q \in \mathbb{Q} \cap [0,1]$, suppose $f_r(x) > g_q(x)$ for some $x \in E$. Then, it must be the case that $x \in V_r$ (otherwise $f_r(x) = 0$), $x \notin \bar{V_q}$ (otherwise $g_q(x) = 1$, and

$$r = f_r(x) > g_q(x) = q.$$

However, $q < r$ implies that $V_r \subset \bar{V_r} \subset V_q \subset \bar{V_q}$, so that, if $x \in V_r$, then x must also be contained in $\bar{V_q}$. This is clearly a contradiction, so $f_r(x) \leq g_q(x)$ for any $x \in E$, which implies that

$$f = \sup_{r \in \mathbb{Q} \cap [0,1]} f_r \leq \inf_{r \in \mathbb{Q} \cap [0,1]} g_q = g.$$

To see the reverse inequality, suppose that $f(x) < g(x)$ for some $x \in E$. Then, because $\mathbb{Q}$ is dense in $\mathbb{R}$, there exist $r, q \in \mathbb{Q}$ such that $f(x) < r < q < g(x)$, and because $f(x), g(x) \in [0,1]$, it must be the case that $r, q \in \mathbb{Q} \cap [0,1]$.
Since $f_r(x) \leq f(x) < r$, it must be the case that $x \notin V_r$, while $q < g(x) \leq g_q(x)$ implies that $x \in \bar{V_q}$. However, because $r < q$, by design we have $\bar{V_q} \subset V_r$, so that $x \in \bar{V_q}$ implies $x \in V_r$, a contradiction. It follows that $f(x) = g(x)$ for any $x \in E$, and the proof is complete.
Q.E.D.

Any such f as in the theorem above is called a Urysohn function.

We conclude by proving an important corollary to Urysohn's lemma, the familiar partition of unity. The theorem can be viewed as a generalization of Urysohn's lemma in the case where a compact set is covered not by a single open set, but by a finite union of open sets.

Theorem 1.32 (Partition of Unity)
Let (E,τ) be a locally compact Hausdorff space. Let $V_1,\cdots,V_n \in \tau$ be open sets and K a compact set such that

$$K \subset V_1 \cup \cdots \cup V_n.$$

Then, there exist functions $h_1,\cdots,h_n \in C_c(E,\tau)$ such that $h_i \prec V_i$ for any $1 \leq i \leq n$ and

$$K \prec \sum_{i=1}^{n} h_i.$$

Proof) If $n=1$, this is just Urysohn's lemma, so we assume $n \geq 2$.

For any $x \in K$, $x \in V_{i(x)}$ for some $1 \leq i(x) \leq n$. Because $\{x\}$ is a compact subset contained in the open set $V_{i(x)}$, by theorem 1.14 there exists an open set $U_x \in \tau$ with compact closure such that

$$x \in U_x \subset \bar{U_x} \subset V_{i(x)}.$$

The collection $\{U_x\}_{x\in K}$ is an open cover of K, and as such, by compactness there exist $x_1,\cdots,x_m \in K$ such that

$$K \subset U_{x_1} \cup \cdots \cup U_{x_m}.$$

For any $1 \leq i \leq n$, define $\iota_i = \{1 \leq j \leq m \mid i(x_j) = i\}$; if $\iota_i \neq \emptyset$, then put

$$H_i = \begin{cases} \bigcup_{j\in\iota_i} \bar{U_{x_j}} & \text{if } \iota_i \neq \emptyset \\ \emptyset & \text{if } \iota_i = \emptyset. \end{cases}$$

Since $\bar{U_{x_j}} \subset V_i$ for any $j \in \iota_i$ because $i(x_j) = i$, it follows that H_i is a compact set contained in V_i (this holds even if $H_i = \emptyset$, since the empty set is compact and contained in any set).

Therefore, for any $1 \leq i \leq n$, by Urysohn's lemma there exists a function $g_i \in C_c(E,\tau)$ such that

$$H_i \prec g_i \prec V_i.$$

Define $h_1, \cdots, h_n$ as $h_1 = g_1$ and

$$h_i = \left(\prod_{j=1}^{i-1} (1 - g_j) \right) g_i$$

for $2 \leq i \leq n$. By design, $h_1 = g_1 \prec V_1$.
For any $2 \leq i \leq n$, because h_i is the product of continuous functions, it is also continuous relative to τ. Furthermore,

$$\left(\bigcup_{j=1}^{i-1} \{g_j = 1\} \right) \cup \{g_i = 0\} \subset \{h_i = 0\},$$

it stands to reason that

$$\{h_i \neq 0\} \subset \left(\bigcap_{j=1}^{i-1} \{g_j \neq 1\} \right) \cap \{g_i \neq 0\} \subset \{g_i \neq 0\};$$

because g_i has compact support, and $\overline{\{h_i \neq 0\}} \subset \overline{\{g_i \neq 0\}}$, h_i also has compact support, meaning that $h_i \in C_c(E, \tau)$.
For any $x \in E$, because $g_1(x), \cdots, g_i(x) \in [0,1]$,

$$h_i(x) = \left(\prod_{j=1}^{i-1} (1 - g_j(x)) \right) g_i(x) \in [0,1],$$

and because $\overline{\{g_i \neq 0\}}$ lies in V_i, we have

$$\overline{\{h_i \neq 0\}} \subset \overline{\{g_i \neq 0\}} \subset V_i.$$

As such, by definition, $h_i \prec V_i$.

It remains to show that $K \prec \sum_{i=1} 6nh_i$.
We first derive an expression for the sum in terms of $g_1, \cdots, g_n$ by induction on n.
Suppose $n = 2$. Then, $h_1 + h_2 = g_1 + (1 - g_1)g_2 = 1 - (1 - g_1)(1 - g_2)$.
Now suppose that

$$h_1 + \cdots + h_k = 1 - \prod_{i=1}^{k} (1 - g_i)$$

for some $2 \leq k \leq n - 1$. Then,

$$\begin{aligned} h_1 + \cdots + h_{k+1} = 1 - \prod_{i=1}^{k} (1 - g_i) + h_{k+1} &= 1 - \prod_{i=1}^{k} (1 - g_i) + \left(\prod_{i=1}^{k} (1 - g_i) \right) g_{k+1} \\ &= 1 - \prod_{i=1}^{k+1} (1 - g_i). \end{aligned}$$

By induction,

$$h_1 + \cdots + h_n = 1 - \prod_{i=1}^{n}(1 - g_i).$$

Being the sum of functions in $C_c(E,\tau)$, $\sum_{i=1}^{n} h_i \in C_c(E,\tau)$ because $C_c(E,\tau)$ is a vector space.
For any $x \in E$, since $g_i(x) \in [0,1]$ and $1 - g_i(x) \in [0,1]$ because g_i is a Urysohn function, so that

$$\prod_{i=1}^{n}(1 - g_i(x)) \in [0,1]$$

and thus

$$h_1(x) + \cdots + h_n(x) = 1 - \prod_{i=1}^{n}(1 - g_i(x)) \in [0,1]$$

as well.
For any $x \in K$, because

$$K \subset U_{x_1} \cup \cdots \cup U_{x_m},$$

$x \in U_{x_j}$ for some $1 \le j \le m$. Letting $i = i(x_j)$, it follows that

$$x \in U_{x_j} \subset H_i \prec g_i \prec V_i,$$

so that $g_i(x) = 1$. By implication,

$$h_1(x) + \cdots + h_n(x) = 1 - \prod_{l=1}^{n}(1 - g_l(x)) = 1.$$

Therefore, $\sum_{i=1}^{n} h_i$ is a function in $C_c(E,\tau)$ taking values in $[0,1]$ such that $h_1(x) + \cdots + h_n(x) = 1$ for any $x \in K$. By definition, $K \prec \sum_{i=1}^{n} h_i$ and the proof is complete.
Q.E.D.

The collection $\{h_1, \cdots, h_n\}$ constructed above is called a partition of unity on K with respect to the cover $\{V_1, \cdots, V_n\}$.

1.14 Urysohn's Lemma for Normal Spaces

Urysohn's lemma in the previous section was formulated in terms of an open set and a compact set contained in that open set. It is also possible to obtain a formulation of Urysohn's lemma that relaxes the compactness assumption by strengthening the separation axiom of the underlying space.

So far, we have worked under the Hausdorff T^2 separation axiom: recall that a topological space (E,τ) is Hausdorff if, for any distinct points $x,y \in E$, there exist disjoint open sets $A,B \in \tau$ such that $x \in A$ and $y \in B$. This is called a separation axiom because it posits how two distinct points can be separated by open sets.

We can also work under two other separation axioms, the T^3 and T^4 axioms. A topological space (E,τ) in which one-point sets are closed is said to be:

- **Regular, or** T^3,
  if for any point $x \in E$ and a closed set B such that $x \notin B$, there exist disjoint open sets $A_1, A_2 \in \tau$ such that $x \in A_1$ and $B \subset A_2$.

- **Normal, or** T^4,
  if for any two disjoint closed sets B_1, B_2, there exist disjoint open sets $A_1, A_2 \in \tau$ such that $B_i \subset A_i$ for $i = 1,2$.

Simply put, the T^2 axiom deals with the separation of points, the T^3 axiom with the separation of a point and a closed set, and the T^4 axiom with the separation of closed sets. The numbers appended to the separation axioms represent how general each axiom is; under the assumption that one-point sets are closed, we can see that a normal space is regular, and that a regular space is Hausdorff. For this reason, we sometimes call regular (normal) spaces regular Hausdorff (normal Hausdorff) spaces.

The following furnishes simple characterizations of regularity and normality similar to theorem 1.14:

Theorem 1.33 (Characterization of Regularity and Normality)
Let (E,τ) be a topological space where one-point sets are closed. Then,

i) E is regular if and only if, for any $x \in E$ and a neighborhood $U \in \tau$ of x, there exists a neighborhood $V \in \tau$ of x such that $x \in V \subset \overline{V} \subset U$.

ii) E is normal if and only if, for any closed set A and an open set U such that $A \subset U$, there exists an open set V such that $A \subset V \subset \overline{V} \subset U$.

Proof) We prove the second claim; the first claim follows by replacing the closed set in question with a single point $x \in E$, since one-point sets are closed by assumption. Suppose (E,τ) is a normal space, and let A be a closed set that is contained in some open set U. In this case, since A and U^c are disjoint closed sets, there exist disjoint open sets W,V such that $A \subset W$ and $U^c \subset V$ by the normality property. Since W,V are disjoint, so is $\overline{W}$ and V; to see this, let $x \in \overline{W}$, and suppose that $x \in V$ as well. Then, because V is a neighborhood of x, the characterization of the closure in lemma 1.12 tells us that $V \cap W \neq \emptyset$, a contradiction. It follwos that $\overline{W} \cap V = \emptyset$. This, together with the fact that $V^c \subset U$, shows us that

$$A \subset W \subset \overline{W} \subset V^c \subset U.$$

Conversely, suppose that, for any closed set A and an open set U containing A, there exists an open set V such that $A \subset V \subset \overline{V} \subset U$. Then, for any disjoint closed subsets B_1, B_2 of E, B_2^c is an open set containing the closed set B_1, and as such there exists an open set V such that

$$B_1 \subset V \subset \overline{V} \subset B_2^c.$$

V and $\overline{V}^c$ are disjoint open sets such that $B_1 \subset V$ and $B_2 \subset \overline{V}^c$, so (E, τ) is normal by definition.

Q.E.D.

Normal spaces are particularly useful, since most of the spaces that we work with are normal, including metric spaces. This is shown in the following result:

Theorem 1.34 (Ubiquity of Normal Spaces)
The following hold true:

i) Every second countable regular space is normal.

ii) Every compact Hausdorff space is normal.

iii) Every metrizable space is normal.

Proof) i) Let (E, τ) be a second countable regular space, and $\mathbb{B} \subset \tau$ the countable base on E that generates τ. Choose any two disjoint closed sets A, B. For any $x \in A$, the disjointness of A and B implies that $x \notin B$, and by regularity, there exists some $W_x \in \tau$ such that $x \in W_x$ and $W_x \cap B = \emptyset$. By the characterization of regularity shown above, there exists an open set $T_x \in \tau$ such that $x \in T_x \subset \overline{T_x} \subset W_x$. Finally, by the definition of a generating base, we can choose some $U_x \in \mathbb{B}$ such that

$$x \in U_x \subset T_x \subset \overline{T_x} \subset W_x \subset B^c.$$

The collection $\{U_x\}_{x \in A}$ of open sets covers A, and because $\mathbb{B}$ is countable, we can index the collection using the natural numbers as $\{U_n\}_{n \in N_+}$. Note that the union $\bigcup_n U_n$ is an open set that does not intersect B, since each U_n is contained in B^c.

We can similarly construct a countable open covering $\{V_n\}_{n\in N_+}$ of B whose union $\bigcup_n V_n$ is an open set that does not intersect A. It remains to restrict each U_n and V_n so that their respective unions do not intersect one another. To this end, for any $n \in N_+$ define

$$U_n' = U_n \setminus \left(\bigcup_{i=1}^{n} \overline{V_i} \right) \quad \text{and} \quad V_n' = V_n \setminus \left(\bigcup_{i=1}^{n} \overline{U_i} \right),$$

and let

$$U = \bigcup_n U_n' \quad \text{and} \quad V = \bigcup_n V_n'.$$

We can immediately see that each U_n' and V_n' are open sets because they involve the intersection of an open set and a finite intersection of open sets (which is also open); by implication, U and V are also open. Furthermore, because $\overline{U_i} \subset B^c$ for each $i \in N_+$, it follows that, for any $n \in N_+$,

$$B \subset \bigcap_{i=1}^{n} \overline{U_i}^c,$$

which implies that

$$B \subset V_n \cap \left(\bigcap_{i=1}^{n} \overline{U_i}^c \right) = V_n'.$$

Therefore, B is contained in V, and likewise, A is conatined in U. Finally, suppose that $x \in U \cap V$. Then, there exists some $n, m \in N_+$ such that $x \in U_n'$ and $x \in V_m'$. Letting $n \geq m$ without loss of generality, note that this implies

$$x \in \bigcap_{i=1}^{n} \overline{V_i}^c \quad \text{and} \quad x \in V_m,$$

and in particular that $x \notin \overline{V_m}$ and $x \in V_m$ at the same time. This is a contradiction, so it must be the case that U and V are disjoint. We have thus shown that any two disjoint closed sets must be contained in disjoint open sets, or that the space (E, τ) is normal.

ii) Let (E, τ) be a compact Hausdorff space. We show first that (E, τ) is regular. Let B be a closed subset of E and x a point that is not contained in B. Since the closed subsets of a compact set is also compact, B is a compact set. For any $y \in B$, since $x \neq y$, the Hausdorff property implies that there exists a neighborhood A_y of y such that $x \notin \overline{A_y}$. The collection $\{A_y\}_{y\in A}$ forms an open cover of B, and by the compactness of B, there exists a finite subcollection $A_{y_1}, \cdots, A_{y_n}$ of open sets such that

$$B \subset A_{y_1} \cup \cdots \cup A_{y_n}.$$

Defining

$$U = A_{y_1} \cup \cdots \cup A_{y_n} \quad \text{and} \quad V = \overline{A_{y_1}}^c \cap \cdots \cap \overline{A_{y_n}}^c,$$

U and V are disjoint open sets such that U contains B and V contains x. This shows us that (E,τ) is regular.

We can prove in a similar manner that (E,τ) is normal. Choose any disjoint closed sets A,B; for any $x \in A$, by regularity there exists an open set V_x containing x such that $\overline{V_x} \cap B = \emptyset$. $\{V_x\}_{x\in A}$ is an open cover of A, and by the compactness of A, there exists a finite number of points $x_1,\cdots,x_n \in A$ such that

$$A \subset V_{x_1} \cup \cdots \cup V_{x_n}.$$

Then, defining $V = V_{x_1} \cup \cdots \cup V_{x_n}$ and $U = \overline{V_{x_1}}^c \cap \cdots \cap \overline{V_{x_n}}^c$, U and V are disjoint open sets such that V contains A and U contains B. This shows us that (E,τ) is normal. Notice in both cases how the finite subcover furnished by compactness allows us to construct an open set via taking a finite intersection of open sets.

iii) Let (E,τ) be a metrizable topological space, and let d be the metric that induces τ. Choose any disjoint closed subsets A,B of E. For any $x \in A$, because x is a point in the open set B^c, there exists some $\epsilon_x > 0$ such that $B_d(x,\epsilon_x) \subset B^c$. Likewise, for any $y \in B$, since $y \in A^c$ where A^c is open, there exists a $\delta_y > 0$ such that $B_d(y,\delta_y) \subset A^c$. Define

$$U = \bigcup_{x\in A} B_d(x,\epsilon_x/2) \quad \text{and} \quad V = \bigcup_{y\in B} B_d(y,\delta_y/2).$$

Both U and V are open, and they contain A and B, respectively. It remains to show that U and V are disjoint. To this end, suppose $z \in U \cap V$. Then, there exist $x \in A$ and $y \in B$ such that $z \in B_d(x,\epsilon_x/2)$ and $z \in B_d(y,\delta_y/2)$. Suppose without loss of generality that $\epsilon_x \leq \delta_y$. Then,

$$d(x,y) \leq d(x,z) + d(y,z) < \frac{\epsilon_x}{2} + \frac{\delta_y}{2} \leq \delta_y,$$

so that $x \in B_d(y,\delta_y)$. Thus, $x \in B_d(y,\delta_y) \cap A$ and $B_d(y,\delta_y) \cap A \neq \emptyset$, which contradicts the fact that $B_d(y,\delta_y) \subset A^c$. As such, U and V are disjoint, and the proof is complete.

Q.E.D.

Note that one-point sets are closed in all three cases: for regular spaces, compact spaces and metrizable spaces.

We are now ready to present the more general version of Urysohn's lemma for normal spaces. This is usually the theorem that is referred to as Urysohn's lemma.

Theorem 1.35 (Urysohn's Lemma)
Let (E,τ) be a normal space. For any closed set F and open set V such that $F \subset V$, there exists a continuous function $f : E \to \mathbb{R}$ such that

$$f(x) \in \begin{cases} \{0\} & \text{if } x \notin V \\ [0,1] & \text{if } x \in V \setminus F \\ \{1\} & \text{if } x \in F \end{cases}$$

for any $x \in E$. This function is called a Urysohn function, and satisfies $I_F \leq f \leq I_V$.

Proof) The proof mirrors the proof for the locally compact Hausdorff case almost one for one. For the sake of completeness, we re-state the proof here.

First, we obtain a set of decreasing sets indexed by the rationals. Let $\{r_n\}_{n \in N_+}$ be the rationals in $(0,1)$ arranged in a sequence (which is possible due to the countability of $\mathbb{Q}$), and denote $r_{-1} = 1, r_0 = 0$. By the characterization of normality, there exists an open set V_1 such that

$$F \subset V_1 \subset \overline{V_1} \subset V.$$

Since $\overline{V_1}$ is itself a closed subset of V, by the characterization again, there exists an open set V_0 such that

$$F \subset V_1 \subset \overline{V_1} \subset V_0 \subset \overline{V_0} \subset V.$$

Suppose that open sets $V_{r_{-1}}, V_{r_0}, V_{r_1}, \cdots, V_{r_n}$ have been chosen for some $n \geq 0$ so that

$$\overline{V_r} \subset V_q \quad \text{for any } r, q \in \{r_{-1}, \cdots, r_n\} \text{ such that } q < r.$$

Considered the set $\mathbb{Q}_n = \{r_{-1}, \cdots, r_n, r_{n+1}\}$ of rationals in $[0,1]$. Then, r_{n+1} must cannot be the maximal or minimal element of this set, since 0 and 1 are both contained in it. It follows that there exist $-1 \leq i, j \leq n$ such that $r_i < r_{n+1} < r_j$, and no rational number in $\mathbb{Q}_n$ lies between r_i and r_{n+1}, nor between r_{n+1} and r_j. Since $\overline{V_j} \subset V_i$, by the characterization of normality again, there exists an open set V_{n+1} such that

$$\overline{V_j} \subset V_{n+1} \subset \overline{V_{n+1}} \subset V_i.$$

It follows that, once we have chosen V_{n+1} in this manner, $\overline{V_r} \subset V_q$ for any $r, q \in \{r_{-1}, \cdots, r_{n+1}\}$ such that $q < r$.

Continuing on in this fashion, we obtain a sequence $\{V_r\}_{r \in \mathbb{Q} \cap [0,1]}$ of open sets such that

$$\overline{V_r} \subset V_q \quad \text{if } q < r$$

and $F \subset V_1 \subset \overline{V_1} \subset V_0 \subset \overline{V_0} \subset V$.

Now we define the sequences $\{f_r\}_{r \in \mathbb{Q} \cap [0,1]}$ and $\{g_r\}_{r \in \mathbb{Q} \cap [0,1]}$ of functions on E as follows:

$$f_r(x) = \begin{cases} r & \text{if } x \in V_r \\ 0 & \text{if } x \notin V_r \end{cases}, \quad g_r(x) = \begin{cases} 1 & \text{if } x \in \overline{V_r} \\ r & \text{if } x \notin \overline{V_r} \end{cases}$$

for any $r \in \mathbb{Q} \cap [0,1]$ and $x \in E$. These functions are both bounded below by 0 and above by 1, and each f_r (g_r) is lower (upper) semicontinuous. To see this, note that

$$\{x \in E \mid a < f_r(x)\} = \begin{cases} E & \text{if } a < 0 \\ V_r & \text{if } 0 \leq a < r \\ \emptyset & \text{if } r \leq a \end{cases}$$

$$\{x \in E \mid g_r(x) < a\} = \begin{cases} E & \text{if } 1 < a \\ \overline{V_r}^c & \text{if } r < a \leq 1 \\ \emptyset & \text{if } a \leq r \end{cases}$$

are both open sets regardless of the value of $a \in \mathbb{R}$.

Heuristically, we may view $\{f_r\}_{r \in \mathbb{Q} \cap [0,1]}$ as a sequence of functions that rises closer to 1 as its support becomes narrower, and $\{g_r\}_{r \in \mathbb{Q} \cap [0,1]}$ as a sequence of functions whose peak and trough move away from each other as the area on which they equal 1 widens. Since the area on which these functions equal 1 cannot be larger than V and cannot be narrower than F, a little bit of visualization shows us how

$$f = \sup_{r \in \mathbb{Q} \cap [0,1]} f_r \quad \text{and} \quad g = \inf_{r \in \mathbb{Q} \cap [0,1]} g_r$$

can serve as the candidates for our Urysohn function. The values of f and g are both contained in the set $[0,1]$, and the lower semicontinuity of $\{f_r\}$ and the upper semicontinuity of $\{g_r\}$ imply that f and g are lower and upper semicontinuous, respectively. Furthermore, for any $x \in F$, since $x \in V_r$ and thus $f_r(x) = r$ for any $r \in \mathbb{Q} \cap [0,1]$, we have

$$f(x) = \sup_{r\in\mathbb{Q}\cap[0,1]} f_r(x) = \sup_{r\in\mathbb{Q}\cap[0,1]} r = 1.$$

Likewise, for any $x \notin V$, since $x \notin \overline{V_r}$ and thus $f_r(x) = 0$ for any $r \in \mathbb{Q}\cap[0,1]$, we have

$$f(x) = \sup_{r\in\mathbb{Q}\cap[0,1]} f_r(x) = 0.$$

This shows us that f takes values in $\{1\}$ on F, $\{0\}$ on V, and $[0,1]$ on $V\setminus F$. It remains to show that f is continuous, which can be done easily if we just show that $f = g$, since in this case f would be both lower and upper semicontinuous.

We first show that $f_r \leq g_q$ for any $r, q \in \mathbb{Q}\cap[0,1]$. Suppose that the contrary holds, so that there exist $r, q \in \mathbb{Q}\cap[0,1]$ and $x \in E$ such that $f_r(x) > g_q(x)$. In this case, $f_r(x)$ cannot be equal to 0, so that $x \in V_r$, and $g_q(x)$ cannot be equal to 1, so that $x \notin \overline{V_q}$. Therefore, we must have

$$r = f_r(x) > g_q(x) = q,$$

which implies by our choice of the open sets V_r and V_q that $\overline{V_r} \subset V_q$. This implies that $x \in V_q$ but $x \notin \overline{V_q}$, a contradiction, so it follows that $f_r \leq g_q$ for any $r, q \in \mathbb{Q}\cap[0,1]$, and thus that

$$f = \sup_{r\in\mathbb{Q}\cap[0,1]} f_r \leq \inf_{q\in\mathbb{Q}\cap[0,1]} g_q = g$$

on E.

To see that the reverse inequality holds, suppose that there exists some $x \in E$ such that $f(x) < g(x)$. In this case, we can choose rational numbers $r < q$ such that

$$0 \leq f(x) < r < q < g(x) \leq 1.$$

By the definition of f and g,

$$f_r(x) \leq f(x) < r \quad \text{and} \quad q < g(x) \leq g_q(x),$$

so we must have $x \notin V_r$ and $x \in \overline{V_q}$. However, $r < q$ implies $\overline{V_q} \subset V_r$, so that $x \in V_r$, a contradiction. It follows that $f(x) = g(x)$ for any $x \in E$, completing the proof.

Q.E.D.

1.15 Urysohn's Lemma for Metric Spaces

We can formulate an even stronger version of Urysohn's lemma if we further restrict the underlying space to a metric space. Of central importance is the distance function, which we will define shortly.

Let (E,d) be a metric space. For any $x \in E$ and non-empty $A \subset E$, we can define the distance from the point x to the set A as

$$d(x,A) = \inf_{y \in A} d(x,y).$$

Note that the infimum is well-defined and exists in $[0,+\infty)$ by the least upper bound property of the real line because the set $\{d(x,y) \mid y \in A\}$ is non-empty and bounded below by 0. The distance function associated with the set A is the mapping $d(\cdot,A) : E \to [0,+\infty)$. Below are the main properties of the distance function:

Lemma 1.36 Let (E,d) be a metric space and A a non-empty subset of E. Then, the following hold true:

i) $d(x,A) = 0$ if and only if $x \in \overline{A}$.

ii) The distance function $d(\cdot,A)$ is continuous on E.

Proof) i) Let A be a closed set. Suppose $d(x,A) = 0$ for some $x \in A$. Then, for any $n \in N_+$, because $0 = d(x,A) < \frac{1}{n}$, there exists an $x_n \in A$ such that

$$0 \le d(x,x_n) < \frac{1}{n}.$$

The sequence $\{x_n\}_{n \in N_+}$ is contained in A and, taking $n \to \infty$ above, we can see that it converges to x. This means that x is a limit point of A, that is, x is contained in the closure $\overline{A}$ of A.

Conversely, suppose that $x \in \overline{A}$, and assume $d(x,A) > 0$. Then, letting $n \in N_+$ be chosen so that $\frac{1}{n} < d(x,A)$, because x is in the closure of A,

$$B_d(x,1/n) \cap A \neq \emptyset.$$

Choosing $y \in B_d(x,1/n) \cap A$, y is a point in A such that $d(x,y) < \frac{1}{n} < d(x,A)$. This contradicts the definition of $d(x,A)$ as the infimum of $\{d(x,y) \mid y \in A\}$, so $d(x,A)$ must be equal to 0.

ii) Choose any $x,y \in E$. Assume without loss of generality that $d(x,A) \ge d(y,A)$. For any $n \in N_+$, since $d(y,A) < d(y,A) + \frac{1}{n}$, by the definition of the infimum there exists a $z_n \in A$ such that $d(y,A) \le d(y,z_n) < d(y,A) + \frac{1}{n}$. For this z_n, we have

$d(x,A) \le d(x,z_n)$, since the infimum is itself a lower bound. Therefore,

$$|d(x,A) - d(y,A)| = d(x,A) - d(y,A) \le d(x,z_n) - d(y,z_n) + \frac{1}{n} \le d(x,y) + \frac{1}{n},$$

where the last inequality follows from the triangle inequality. Taking $n \to \infty$ on both sides now shows us that

$$|d(x,A) - d(y,A)| \le d(x,y).$$

This holds for arbitrary $x, y \in E$, so we have actually proven the stronger result that $d(\cdot, A)$ is Lipschitz continuous on E with Lipschitz constant 1.

Q.E.D.

We can now present the strongest version of Urysohn's lemma. In the theorem, we use the distance function to directly construct the necessary Urysohn function.

Theorem 1.37 (Urysohn's Lemma for Metric Spaces)
Let (E,d) be a metric space. For any closed set F and open set V such that $F \subset V$, therre exists a continuous function $f : E \to \mathbb{R}$ such that

$$f(x) \in \begin{cases} \{0\} & \text{if } x \notin V \\ [0,1] & \text{if } x \in V \setminus F \\ \{1\} & \text{if } x \in F \end{cases}$$

for any $x \in E$.

In addition, if the distance between F and V^c is bounded below by a non-zero value, that is, if there exists a $\delta > 0$ such that $d(x,y) \ge \delta$ for any $x \in F$ and $y \in V^c$, then we can choose f to be Lipschitz continuous.

Proof) We define $f : E \to \mathbb{R}$ as

$$f(x) = \frac{d(x,V^c)}{d(x,V^c) + d(x,F)}$$

for any $x \in E$. For any $x \in E$, if $d(x,V^c) = 0$ then $x \in V^c$ due to the closedness of V^c, which implies that $x \notin F$ and thus $d(x,F) > 0$. Conversely, if $d(x,F) = 0$ then $x \in F$ and $x \notin V^c$, so that $d(x,V^c) > 0$. As such, $d(x,V^c) + d(x,F) > 0$ for any choice of $x \in E$, which shows us that f is well-defined.

f clearly takes values in $[0,1]$. If $x \in F$, then $d(x,F) = 0$ and $f(x) = 1$. On the other hand, if $x \in V^c$, then $d(x,V^c) = 0$ and $f(x) = 0$. Finally, because the distance functions involved are continuous, and continuity is preserved across arithmetic operations, f is itself continuous. We have thus shown that f is our desired Urysohn function.

Now suppose that there exists a $\delta > 0$ such that $d(x,y) \geq \delta$ for any $x \in F$ and $y \in V^c$. For any $x,y \in E$, assume without loss of generality that $f(x) \geq f(y)$. Then,

$$\begin{aligned}
|f(x) - f(y)| &= \left| \frac{d(x,V^c)}{d(x,V^c) + d(x,F)} - \frac{d(y,V^c)}{d(y,V^c) + d(y,F)} \right| \\
&= \left| \frac{d(x,V^c)(d(y,V^c) + d(y,F)) - d(y,V^c)(d(x,V^c) + d(x,F))}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} \right| \\
&= \left| \frac{d(x,V^c)d(y,F) - d(y,V^c)d(x,F)}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} \right| \\
&= \left| \frac{d(x,V^c)d(y,F) - d(x,V^c)d(x,F) + d(x,V^c)d(x,F) - d(y,V^c)d(x,F)}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} \right| \\
&\leq \frac{d(x,V^c)|d(y,F) - d(x,F)|}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} + \frac{d(x,F)|d(x,V^c) - d(y,V^c)|}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} \\
&\leq \frac{(d(x,V^c) + d(x,F))d(x,y)}{(d(x,V^c) + d(x,F))(d(y,V^c) + d(y,F))} \\
&= \frac{d(x,y)}{d(y,V^c) + d(y,F)},
\end{aligned}$$

where we used the fact that $|d(x,F) - d(y,F)| \leq d(x,y)$ due to the Lipschitz continuity of the distance function.

Suppose that $d(y,V^c) + d(y,F) < \delta$. Then, $d(y,V^c) < \delta - d(y,F)$, so there exists a $z_1 \in V^c$ such that $d(y,z_1) < \delta - d(y,F)$. This in turn implies that $d(y,F) < \delta - d(y,z_1)$, so there exists a $z_2 \in F$ such that $d(y,z_2) < \delta - d(y,z_1)$. By the triangle inequality, we now have

$$d(z_1,z_2) \leq d(y,z_1) + d(y,z_2) < \delta.$$

Since $z_1 \in V^c$ and $z_2 \in F$, this is a contradiction, so it must be the case that $d(y,V^c) + d(y,F) \geq \delta$. By implication,

$$|f(x) - f(y)| \leq \frac{d(x,y)}{d(y,V^c) + d(y,F)} \leq \frac{1}{\delta} d(x,y),$$

and since this holds for any $x,y \in E$, f is Lipschitz continuous with Lipschitz constant equal to $\frac{1}{\delta}$.

Q.E.D.

1.16 The Axiom of Choice and Tychonoff's Theorem

For our last topic, we study one of the most fundamental axioms in mathematics, the axiom of choice, and show how it is equivalent to Zorn's lemma and the Hausdorff maximality principle. We conclude by showing that these imply Tychonoff's theorem, which states that the Cartesian product of compact sets is also compact.

We first state the axiom of choice. Let X be some set and $\mathcal{F}$ a collection of subsets of X. The axiom of choice posits that there exists a choice function $f : \mathcal{F} \to X$ such that $f(S) \in S$ for any $S \in \mathcal{F}$. In other words, given any collection of sets, we can formulate a rule to choose an element from each set.

Given the axiom of choice, we can prove two deep results in mathematics, Zorn's lemma and the Hausdorff Maximality Principle. First, some notation. A set X is said to be partially ordered by the order relation $\leq$ if

- **(Reflexivity)** For any $x \in X$, $x \leq x$
- **(Antisymmetry)** For any $x, y \in X$ such that $x \leq y$ and $y \leq x$, we have $x = y$
- **(Transitivity)** For any $x, y, z \in X$, if $x \leq y$ and $y \leq z$, then $x \leq z$

Note that not all elements of X may be comparable. If every element of a partially ordered set X is comparable, then we say that it is totally ordered by the order relation $\leq$. Clearly, any collection $\mathcal{F}$ of subsets of some set X is partially ordred by the set inclusion operation $\subset$; any subcollection of $\mathcal{F}$ that is totally ordered by $\subset$ is called a subchain of $\mathcal{F}$. The union of any subchain of $\mathcal{F}$ refers to the union of all elements contained in that subchain.

Let X be a partially ordered set with the order relation $\leq$. We say that $m \in X$ is a maximal element of X if, for any $x \in X$, $m \leq x$ implies $m = x$. In other words, there is no element of X that is comparable to m and larger than m. Since a maximal element of X need not be comparable with every element of X, a partially ordered set X may admit more than one maximal element. Clearly, if X is totally ordered, then it has at most one maximal element. A related but distinct concept is that of an upper bound of a set; a subset E of X has upper bound $x \in X$ if $y \leq x$ for any $y \in E$. Note that an upper bound of a set must be comparable with every element of that set, while a maximal element need not be.

The final concept we need is that of the maximal totally ordered subset. A subset E of X is said to be a maximal totally ordered subset of X if it is a totally ordered subset of X that, if expanded in any way, ceases to be totally ordered. Formally, the totally ordered subset E of X is maximal if, for any $x \in X$ such that $x \notin E$, $E \cup \{x\}$ is not totally ordered.

We can now state Zorn's lemma and the Hausdorff Maximality Principle:

- **(Zorn's Lemma)** Let X be a partially ordered set, and assume that every totally ordered subset of X has an upper bound. Then, X has at least one maximal element.
- **(Hausdorff's Maximality Principle)** Let X be a partially ordered set. Then, it contains a maximal totally ordered set.

While Zorn's lemma deals with maximal elements and the maximality principle with maximal subsets, they are in fact equivalent, as we now show:

Theorem 1.38 Zorn's lemma and Hausdorff's Maximality Principle are equivalent.

Proof) Suppose that Zorn's lemma holds, and let X be a partially ordered set under the order relation $\leq$. Define $C(X)$ as the collection of all totally ordered subsets of X; then, $C(X)$ is partially ordered under set inclusion. Let a subcollection $\mathcal{F}$ of $C(X)$ be totally ordered under set inclusion, and define $M = \bigcup_{A \in \mathcal{F}} A$. Choose any $x, y \in M$; letting $x \in A$ and $y \in B$ for $A, B \in \mathcal{F}$, since $\mathcal{F}$ is totally ordered, we can assume without loss of generality that $A \subset B$. Therefore, $x, y \in B$, and because B is a totally ordered subset of X, either $x \leq y$ or $y \leq x$. This shows us that any two elements of M are comparable using $\leq$, or that M is a totally ordered subset of X; by definition, $M \in C(X)$. In addition, $A \subset M$ for any $A \in \mathcal{F}$, making M an upper bound of $\mathcal{F}$. We have shown that, under set inclusion, any totally ordered subset of $C(X)$ has an upper bound in $C(X)$. By Zorn's lemma, $C(X)$ has a maximal element U. This U is a totally ordered subset of X, and suppose that there exists an $x \in X$ such that $U \cup \{x\}$ is also totally ordered. By implication, $U \cup \{x\} \in C(X)$, $U \subset U \cup \{x\}$ but $U \neq U \cup \{x\}$; this contradicts the fact that U is a maximal element of $C(X)$, so U must be a maximal totally ordered subset of X. This proves Hausdorff's maximality principle.

Conversely, suppose that Hausdorff's maximality principle holds, and let X be a partially ordered set under the order relation $\leq$ such that any totally ordered subset of X has an upper bound. By the maximality principle, X has a maximal totally ordered subset E, which has upper bound m by assumption. We will now show that m is a maximal element of X. Let $x \in X$ satisfy $m \leq x$, and suppose that $m < x$. In this case, $x \notin E$ (m is an upper bound of E), and by the maximality of E, and for any $y \in E$, $y \leq m < x$ by transitivity. This makes $E \cup \{x\}$ a totally ordered subset of E, which contradicts the maximality of E. Therefore, we must have $m = x$, and m is a maximal element of X.

Q.E.D.

The two results above follow from the axiom of choice; we show this by demonstrating how the axiom of choice leads to the maximality principle. First, a lemma:

Lemma 1.39 Let X be a nonempty set and $\mathcal{F}$ a nonempty collection of subsets of X such that the union of every subchain in $\mathcal{F}$ is also contained in $\mathcal{F}$. Let $g : \mathcal{F} \to \mathcal{F}$ be a function such that, for any $A \in \mathcal{F}$, $A \subset g(A)$ and $g(A) \setminus A$ consists of at most one element. Then, there exists an $A \in \mathcal{F}$ such that $g(A) = A$.

Proof) We first introduce the concept of towers. Choose any $A_0 \in \mathcal{F}$; we call a subcollection $\mathcal{F}' \subset \mathcal{F}$ a tower if

- $A_0 \in \mathcal{F}'$
- The union of any subchain in $\mathcal{F}'$ is contained in $\mathcal{F}'$
- $g(\mathcal{F}') \subset \mathcal{F}'$, that is, for any $A \in \mathcal{F}'$, we have $g(A) \in \mathcal{F}'$.

Note that the set of all towers is non-empty, since it contains the entire collection $\mathcal{F}$ itself. We can also see that the collection

$$\mathcal{F}_1 = \{A \in \mathcal{F} \mid A_0 \subset A\}$$

is a tower; the first two conditions are clearly satisfied, and for any $A \in \mathcal{F}_1$, $A_0 \subset A \subset g(A)$, so that $g(A) \in \mathcal{F}_1$ as well.

Define $\mathcal{F}_0$ as the intersection of all towers. $\mathcal{F}_0$ is also a tower because, first, A_0 is contained in every tower and thus in $\mathcal{F}_0$ as well; second, any subchain of $\mathcal{F}_0$ is contained in every tower, so that its union is also contained in every tower and thus in $\mathcal{F}_0$ as well; and third, for any $A \in \mathcal{F}_0$, A is contained in every tower, so that $g(A)$ is also contained in every tower and thus in $\mathcal{F}_0$. In addition, because $\mathcal{F}_0 \subset \mathcal{F}_1$, $A_0 \subset A$ for any $A \in \mathcal{F}_0$.

Suppose $\mathcal{F}_0$ is a subchain of $\mathcal{F}$, that is, it is totally ordered under set inclusion. Seeing as how $\mathcal{F}_0$ is its own subchain in this case, its union A^* will be contained in $\mathcal{F}_0$ and therefore $g(A^*) \in \mathcal{F}_0$ by the properties of a tower. This implies that $g(A^*) \subset A^*$, and because $A^* \subset g(A^*)$ by definition, we have $g(A^*) = A^*$, completing our proof. Thus, we need only show that $\mathcal{F}_0$ is a subchain of $\mathcal{F}$.

To this end, we define Γ as a subcollection of $\mathcal{F}_0$ that collects every set in $\mathcal{F}_0$ that is comparable to any other set in $\mathcal{F}_0$ under set inclusion. If $\Gamma = \mathcal{F}_0$, this means that every set of $\mathcal{F}_0$ is comparable to any other set in $\mathcal{F}_0$, and therefore that $\mathcal{F}_0$ is totally ordered under set inclusion. To establish this equality, we need only show that Γ is a tower, implying that $\mathcal{F}_0 \subset \Gamma$. We thus focus on showing that Γ is a tower.

Γ satisfies the first two conditions of a tower: $A_0 \in \Gamma$ because A_0 is contained in every set in $\mathcal{F}_0$. Furthermore, for any subchain in Γ and some $A \in \mathcal{F}_0$, if there exists at least one set in that subchain that contains A, then its union contains A, while if every set in the subchain is contained in A, then its union is contained in A. In any case, the union of the subchain is comparable with $A \in \mathcal{F}_0$, and since A was chosen arbitrarily, this indicates that the union is contained in Γ.

The third condition is trickier to establish; we must show that, for any $A \in \Gamma$, $g(A) \in \Gamma$ as well. First fix $A \in \Gamma$, and define $\Phi(A)$ as the subcollection of $\mathcal{F}_0$ consisting of sets B such that either $B \subset A$ or $g(A) \subset B$; that is, $\Phi(A)$ is the collection of all sets in $\mathcal{F}_0$ comparable to $g(A)$. If we can show that $\Phi(A)$ is a tower, then $\Phi(A) = \mathcal{F}_0$ and $B \subset A \subset g(A)$ or $g(A) \subset B$ for any $B \in \mathcal{F}_0$; in other words, $g(A)$ is comparable to every set in $\mathcal{F}_0$. This shows us that $g(A) \in \Gamma$, and it follows that Γ is a tower. Therefore, the final step of our proof involves proving that $\Phi(A)$ is a tower.

$\Phi(A)$ satisfies the first two conditions of a tower: $A_0 \in \Phi(A)$ because $A \in \Gamma \subset \mathcal{F}_0$ and thus $A_0 \subset A \subset g(A)$, making A_0 comparable to $g(A)$. For any subchain in $\Phi(A)$, if at least one set in the subchain contains $g(A)$, then its union contains $g(A)$, while if every set in the subchain is contained in A, then its union is also contained in A. This shows us that the union of this subchain is also contained in $\Phi(A)$. Finally, choose any $B \in \Phi(A)$; we must show that $g(B) \in \Phi(A)$ as well. Since either $B \subset A$ or $g(A) \subset B$, we consider the following cases:

- **B is a proper subset of A**
  If $g(B) = B$, then $g(B) \subset A$. If B is a proper subset of $g(B)$, then A cannot be a proper subset of $g(B)$; otherwise $g(B) \setminus B$ contains at least two elements, a contradiction. $A \in \Gamma$, which means that it is comparable to any set in $\mathcal{F}_0$, so we must have $g(B) \subset A$ if A is not a proper subset of $g(B)$. In any case, $g(B) \subset A$.

- $B = A$
  In this case, $g(B) = g(A)$, so that $g(A) \subset g(B)$ trivially.

- $g(A) \subset B$
  In this case, $g(A) \subset g(B)$ trivially, since $B \subset g(B)$.

Thus, we can see that either $g(B) \subset A$ or $g(A) \subset g(B)$. By definition, $g(B) \in \Phi(A)$, and since this holds for any $B \in \Phi(A)$, $\Phi(A)$ is a tower, completing our proof.

Q.E.D.

The lemma allows us to prove the maximality principle in conjunction with the axiom of choice:

Theorem 1.40 (Hausdorff's Maximality Principle)
If the axiom of choice holds, then so does Hausdorff's Maximality Principle.

Proof) Choose any partially ordered set X under the order relation $\leq$. We must show that X contains a maximally totally ordered subset. If X is empty, then the maximality principle is trivially satisfied. Suppose now that X is non-empty. Let $\mathcal{F}$ be the collection of any subset of X that is totally ordered under $\leq$. Then, $\mathcal{F}$ is non-empty because any singleton is trivially totally ordered. Choose any subchain Φ of $\mathcal{F}$, and let A be its union. Then, for any $x, y \in A$, there exist $B_x, B_y \in \Phi$ such that $x \in B_x$ and $y \in B_y$. Since Φ is totally ordered under set inclusion, without loss of generality we can assume that $B_x \subset B_y$, so that $x, y \in B_y$. Finally, since B_y is a totally ordered set by design under $\leq$, x and y are comparable under $\leq$. This shows us that A is totally ordered under $\leq$, so that $A \in \mathcal{F}$. Therefore, the union of any subchain of $\mathcal{F}$ is also contained in $\mathcal{F}$.

Let $f : \mathcal{F} \to X$ be a choice function, that is, a function such that $f(A) \in A$ for any $A \in \mathcal{F}$; such a function exists by the axiom of choice. Now we can define the function $g : \mathcal{F} \to \mathcal{F}$ in the following way: for any $A \in \mathcal{F}$, define

$$A^* = \{x \in X \mid x \in A^c,\ A \cup \{x\} \in \mathcal{F}\};$$

if $A^* = \emptyset$, we put $g(A) = A$, while if $A^* \neq \emptyset$, we put $g(A) = A \cup \{f(A^*)\} \in \mathcal{F}$. Then, g is exactly the type of function described in lemma 2.17, and $\mathcal{F}$ satisfies the conditions of lemma 2.17. as well. As such, by that lemma, there exists a $A \in \mathcal{F}$ such that $g(A) = A$. This indicates that $A^* = \emptyset$, or that there exists no $x \in X$ not contained in A that makes $A \cup \{x\}$ a totally ordered set.

Q.E.D.

Zorn's lemma and the maximality principle have plenty of applications; famously, Zorn's lemma can be used to show that every vector space has a basis.

Theorem 2.19 Let V be a vector space over a field F. Then, V has a basis, that is, a linearly independent subset of V that spans V.

Proof) If V is an empty set, then the empty set itself is a linearly independent collection that spans V, and the proof is completed.

Now suppose that V is non-empty, and let $\mathcal{F}$ be the collection of all linearly independent sets in V. Note that $\mathcal{F}$ is partially ordered by set inclusion. To use Zorn's lemma, we will first show that every totally ordered set in $\mathcal{F}$ has an upper bound. Let $\mathcal{F}_0$ be a subset of $\mathcal{F}$ that is totally ordered by set inclusion, and define B as the subset of V that contains all the elements of $\mathcal{F}_0$. We will show that B is an upper bound of $\mathcal{F}_0$. Clearly, all the sets in $\mathcal{F}_0$ are contained in B, so that B dominates the sets in $\mathcal{F}_0$ under the set inclusion order. It remains to show that $B \in \mathcal{F}$, or that B is a collection of linearly independent vectors. Suppose not. In this case, there exists a finite collection $\{v_1, \cdots, v_k\}$ of linearly dependent vectors in B. Since $v_1, \cdots, v_k$ are elements of B, there exist $S_1, \cdots, S_k \in \mathcal{F}_0$ such that $v_i \in S_i$ for any $1 \leq i \leq k$. $\mathcal{F}_0$ is totally ordered in the set inclusion order, so there exists a $1 \leq j \leq k$ such that $S_i \subset S_j$ for any $1 \leq i \leq k$. This implies that the linearly dependent set $\{v_1, \cdots, v_k\}$ is contained in the linearly independent set S_j, a contradiction. Therefore, it must be the case that $B \in \mathcal{F}_0$, making B an upper bound of $\mathcal{F}_0$ in $\mathcal{F}$.

We can now apply Zorn's lemma to $\mathcal{F}$ to conclude that $\mathcal{F}$ has a maximal element M. It remains to show that M spans V to conclude that M is a basis of V. Suppose not. Letting $v \in V$ be a non-zero vector that is not spanned by M (and thus not contained in M), choose any finite subcollection $\{v_1, \cdots, v_k\}$ of M and assume that there exist scalars $a_1, \cdots, a_k, a \in F$ such that

$$a_1 v_1 + \cdots + a_k v_k + av = \mathbf{0}.$$

If $a \neq 0$, then v is spanned by $\{v_1, \cdots, v_k\}$, contradicting the assumption that v is unspanned by M. Thus, $a = 0$, and since this implies that

$$a_1 v_1 + \cdots + a_k v_k = \mathbf{0},$$

and $v_1, \cdots, v_k$ are linearly independent, we must have $a_1 = \cdots = a_k = 0$ as well. By definition, $\{v_1, \cdots, v_k, v\}$ is a linearly independent set, and since this holds for any finite subcollection $\{v_1, \cdots, v_k\}$ of M, the set $M \cup \{x\}$ is a linearly independent set. However, this contradicts the maximality of M, and as such M is a basis of V.

Q.E.D.

Zorn's lemma can also be used to prove a fundamental theorem in topology stating that the Cartesian product of compact spaces is compact relative to the corresponding product topology. This result is stated below:

Theorem 1.41 (Tychonoff's Theorem)
Let $\{(E_i, \tau_i)\}$ be an arbitrary collection of compact topological spaces. Then, the Cartesian product $E = \prod_i E_i$ is compact in the product topology $\tau = \prod_i \tau_i$.

Proof) We rely on the FIP characterization of compactness to prove that E is compact. That is, we choose some collection $\mathcal{A} = \{A_\alpha\}$ of subsets of E that are closed in τ and possess the finite intersection property, and show that their intersection $\cap_\alpha A_\alpha$ is non-empty.

We first use Zorn's lemma to furnish a maximal collection of subsets of E with the FIP. Define

$$P = \{\mathcal{B} \mid \mathcal{A} \subset \mathcal{B} \subset 2^E,\ \mathcal{B} \text{ has the FIP}\}.$$

Note that P is a partially ordered set under set inclusion, where we treat subsets of E as the "elements" of the collections $\mathcal{B} \in P$. To apply Zorn's lemma, we need only show that every totally ordered subset of P has an upper bound in P. Choose any (nonempty) totally ordered subset $\{\mathcal{B}_i\}$ of P, and define

$$\mathcal{B} = \bigcup_i \mathcal{B}_i,$$

that is, we let $\mathcal{B}$ be the collection of subsets of E that contains every subset of E contained in each $\mathcal{B}_i$. Clearly, each $\mathcal{B}_i \subset \mathcal{B}$. It remains to prove that $\mathcal{B}$ is contained in P, that is, it has the FIP and includes every subset contained in $\mathcal{A}$. The latter point is trivial, since $\mathcal{A} \subset \mathcal{B}_i \subset \mathcal{B}$ for any i. For the former, choose any $A_1, \cdots, A_n \in \mathcal{B}$; then, for any $1 \leq i \leq n$, there must exist an m_i such that $A_i \in \mathcal{B}_{m_i}$. $\{\mathcal{B}_m\}$ is totally ordered under set inclusion, so there must exist some $1 \leq j \leq n$ such that $\mathcal{B}_{m_i} \subset \mathcal{B}_{m_j}$ for any $1 \leq i \leq n$; this implies that $A_1, \cdots, A_n \in \mathcal{B}_{m_j}$, and because $\mathcal{B}_{m_j}$ possesses the FIP, the intersection $A_1 \cap \cdots \cap A_n \neq \emptyset$. As such, $\mathcal{B}$ has the FIP, and is contained in $\mathcal{B}$.

By Zorn's lemma, there exists a maximal element $\mathcal{D} = \{D_\beta\}$ of P, that is, a collection of subsets of E with the FIP such that every set in $\mathcal{A}$ is also contained in $\mathcal{D}$ and any extension of $\mathcal{D}$ fails to have the FIP. Now the claim is proven if the intersection $D = \cap_\beta \overline{D_\beta}$ is non-empty, since $\mathcal{A}$, being a collection of closed sets, is contained in $\{\overline{D_\beta}\}$.

One useful property of the maximality of $\mathcal{D}$ that we use below is that the intersection of any finite colleciton of sets in $\mathcal{D}$ (which is non-empty by the FIP of $\mathcal{D}$) is also contained in $\mathcal{D}$. To see this, choose any $D_1, \cdots, D_n \in \mathcal{D}$ and define

$$A = D_1 \cap \cdots \cap D_n.$$

Consider the extension $\mathcal{D}_1 = \mathcal{D} \cup \{A\}$. By the FIP of $\mathcal{D}$, $\mathcal{D}_1$ also possesses the FIP, and it contains $\mathcal{A}$ because $\mathcal{D}$ does. Therefore, by the maximality of $\mathcal{D}$, A must be a set contained in $\mathcal{D}$.

For any i, let $\pi_i : E \to E_i$ be the natural projection of E onto E_i, that is,

$$\pi_i\left((x_j)_j\right) = x_i$$

for any $(x_j)_j \in E$. This function is clearly continuous relative to the product topology τ and the marginal topology τ_i, since the inverse image of any $A_i \in \tau_i$ with respect to π_i is an open rectangle on E. Note that $\{\overline{\pi_i(D_\beta)}\}_\beta$ is a collection of closed subsets of E_i that has the FIP, since $\mathcal{D}$ has the FIP. Therefore, by the compactness of E_i and the FIP characterization of compactness, there exists an

$$x_i \in \bigcap_\beta \overline{\pi_i(D_\beta)}.$$

We can show that, for any index i and neighborhood $N_i \in \tau_i$ of x_i, the set $\pi_i^{-1}(N_i)$ is contained in the collection $\mathcal{D}$. For any β, since $x_i \in \overline{\pi_i(D_\beta)}$, it follows that $N_i \cap \pi_i(D_\beta) \neq \emptyset$. This means that there exists some $y \in D_\beta$ such that $\pi_i(y) = y_i \in N_i$; this allows us to write $y \in \pi_i^{-1}(N_i)$, so that

$$y \in \pi_i^{-1}(N_i) \cap D_\beta,$$

making $\pi_i^{-1}(N_i) \cap D_\beta \neq \emptyset$. This holds for any β, so $\pi_i^{-1}(N_i)$ intersects every set in $\mathcal{D}$. We can now show that the collection $\mathcal{D}_1 = \mathcal{D} \cup \{\pi_i^{-1}(N_i)\}$ has the FIP. For any finite collection of sets $D_1, \cdots, D_n$ in $\mathcal{D}_1$, if $D_1, \cdots, D_n$ are contained in $\mathcal{D}$, then $D_1 \cap \cdots \cap D_n \neq \emptyset$ by the FIP of $\mathcal{D}$. On the other hand, if, say, $D_1 = \pi_i^{-1}(N_i)$, then

$$D_1 \cap D_2 \cap \cdots \cap D_n = \left(\bigcap_{i=2}^{n} D_i \right) \cap \pi_i^{-1}(N_i).$$

By the maximality of $\mathcal{D}$, $\bigcap_{i=2}^{n} D_i$, being a finite intersection of elements of $\mathcal{D}$, is also contained in $\mathcal{D}$. Furthermore, since $\pi_i^{-1}(N_i)$ intersects every element of $\mathcal{D}$, it follows that $D_1 \cap D_2 \cap \cdots \cap D_n \neq \emptyset$, and $\mathcal{D}_1$ has the FIP. $\mathcal{D}_1$ contains $\mathcal{A}$ because $\mathcal{D}$ does, so by the maximality of $\mathcal{D}$, it follows that

$$\pi_i^{-1}(N_i) \in \mathcal{D}.$$

Finally, defining $x = (x_i)_i \in E$, we shall prove that $x \in D$. Suppose that $x \notin D$. Then, there exists some β such that $x \notin \overline{D_\beta}$, which in turn implies that there exists a neighborhood $N \in \tau$ of x such that $N \cap D_\beta = \emptyset$. The base of all open rectangles generates τ, so there exists some open rectangle $\prod_i N_i$ such that

$$x \in \prod_i N_i \subset N,$$

implying that $\prod_i N_i \cap D_\beta = \emptyset$. Letting $i_1, \cdots, i_n$ be the indices such that $N_i \neq E_i$, we can express the open rectangle $\prod_i N_i$ as

$$\prod_i N_i = \bigcap_{j=1}^{n} \pi_{i_j}^{-1}\left(N_{i_j}\right).$$

Since N_{i_j} is an open neighborhood of x_{i_j}, each $\pi_{i_j}^{-1}\left(N_{i_j}\right) \in \mathcal{D}$ by the result above. As such,

$$\prod_i N_i \cap D_\beta = \left(\bigcap_{j=1}^{n} \pi_{i_j}^{-1}\left(N_{i_j}\right) \right) \cap D_\beta,$$

being a finite intersection of elements of $\mathcal{D}$, must be non-empty, which contradicts the fact that $\prod_i N_i \cap D_\beta = \emptyset$ under our assumption. It follows that

$$x \in D = \bigcap_\beta \overline{D_\beta},$$

and as such, $\bigcap_\alpha A_\alpha \neq \emptyset$.

Q.E.D.

1.17 Sequences and Subsequences

This section deals with the analysis of subsequences, which is, strictly speaking, not a subject of topology. Nevertheless, it is of much use to measure theory, so we introduce some important results. First, note the selection theorem for subsequences, which is a fundamental result we exploit often to prove the convergence of sequences.

Lemma 1.42 (The Selection Theorem)
Let (E,d) be a metric space, and $\{x_n\}_{n\in N_+}$ a sequence in E. $\{x_n\}_{n\in N_+}$ converges to some $x \in E$ if and only if every subsequence of $\{x_n\}_{n\in N_+}$ has a further subsequence that converges to x.

Proof) Necessity follows easily because, if $\{x_n\}_{n\in N_+}$ converges to $x \in E$, then every subsequence of $\{x_n\}_{n\in N_+}$ also converges to x (the "further subsequence" mentioned in the statement of the theorem can be chosen to be the subsequence itself).

As for sufficiency, let every subsequence of $\{x_n\}_{n\in N_+}$ have a further subsequence that converges to $x \in E$, and suppose that $\{x_n\}_{n\in N_+}$ does not itself converge to x. By definition, there exists an $\epsilon > 0$ such that, for any $N \in N_+$, $d(x,x_n) \geq \epsilon$ for some $n \geq N$. Choose $n_1 \in N_+$ so that $d(x,x_{n_1}) \geq \epsilon$. Assuming that $n_1 < \cdots < n_k$ have been chosen for some $k \geq 1$, choose $n_{k+1} \geq n_k + 1$ so that $d(x,x_{n_{k+1}}) \geq \epsilon$. Then, $\{x_{n_k}\}_{k\in N_+}$ is a subsequence of $\{x_n\}_{n\in N_+}$ such that $d(x,x_{n_k}) \geq \epsilon$ for any $k \in N_+$. This indicates that no subsequence of the subsequence $\{x_{n_k}\}_{k\in N_+}$ can converge to x, which is a contradiction. Therefore, $\{x_n\}_{n\in N_+}$ converges to x.

Q.E.D.

Let (E,d) be a metric space, and $\{x_n\}_{n\in N_+}$ a sequence in E. The set of all limits of subsequences of $\{x_n\}_{n\in N_+}$ is called the set of subsequential limits of $\{x_n\}_{n\in N_+}$, and we denote it by S^*. If $S^* = \emptyset$, then the sequence $\{x_n\}_{n\in N_+}$ has no convergent subsequences.

1.17.1 Sequences on the Extended Real Line

Unlike the real line, the extended real line $[-\infty,+\infty]$ does not admit a metric, which makes the study of convergence of sequences taking values in $[-\infty,+\infty]$ difficult. As such, we extend the concept of convergence of real sequences, which is well-defined because the real line is a metric space under the euclidean metric, to define the convergence of sequences taking values in $[-\infty,+\infty]$. This is important because later on we deal with the (pointwise) convergence of so-called non-negative functions, which are functions taking values in $[0,+\infty]$.

First, we define the suprema and infima of subsets of $[-\infty,+\infty]$. For any non-empty set $A \subset [-\infty,+\infty]$, we let $\sup A$ be equal to the supremum of $A \cap \mathbb{R}$ if A has a real valued upper bound, and to $+\infty$ if otherwise. In any case, $\sup A$ is well-defined, which shows us that $\sup A$

exists in $[-\infty, +\infty]$ for any non-empty subset A of the extended real line. This is arguably a stronger version of the least upper bound property of the real line, since subsets of the extended real line need not have a real upper bound to admit a supremum. The infimum of subsets of the extended real line are defined in a similar manner.

Suppose $\{x_n\}_{n\in N_+}$ is a sequence that takes values in $[-\infty, +\infty]$. We say that this sequence converges to some $x \in \mathbb{R}$ if it converges to x in the usual way, that is, if for any $\epsilon > 0$ there exists an $N \in N_+$ such that $|x_n - x| < \epsilon$ for any $n \geq N$. On the other hand, we say that this sequence converges to $+\infty$ $(-\infty)$ if, for any $M \in \mathbb{R}$, there exists an $N \in N_+$ such that $x_n > M$ $(x_n < M)$ for any $n \geq N$. If $\{x_n\}_{n\in N_+}$ is a real valued sequence, we are essentially defining convergence to $+\infty$ $(-\infty)$ as divergence to to $+\infty$ $(-\infty)$.

Lemma 1.43 Any increasing (decreasing) sequence $\{x_n\}_{n\in N_+}$ taking values in the extended real line converges to its supremum (infimum).

Proof) Let $\{x_n\}_{n\in N_+}$ be a monotonically increasing sequence taking values in $[-\infty, +\infty]$, and denote $x = \sup_{n\in N_+} x_n \in [-\infty, +\infty]$. If $x \in \mathbb{R}$, then for any $\epsilon > 0$, there exists an $N \in N_+$ such that

$$x - \epsilon < x_N \leq x,$$

and since $x_N \leq x_n \leq x$ for any $n \geq N$, we can see that

$$|x_n - x| = x - x_n < \epsilon$$

for any $n \geq N$. This shows us that $\{x_n\}_{n\in N_+}$ converges to x.

On the other hand, if $x = +\infty$, then the set $\{x_n \mid n \in N_+\}$ does not admit a real upper bound, meaning that, for any $M \in \mathbb{R}$, there exists an $N \in N_+$ such that $x_N > M$. Since $x_n \geq x_N$ for any $n \geq N$, this shows us that $x_n > M$ for any $n \geq N$, and by definition, $\{x_n\}_{n\in N_+}$ converges to x.

Finally, if $x = -\infty$, then $x_n = -\infty$ for any $n \in N_+$ and $\{x_n\}_{n\in N_+}$ trivially converges to x. In any case, $\{x_n\}_{n\in N_+}$ converges to its supremum x; the proof for monotonically decreasing sequences is almost identical.

Q.E.D.

We say that the sequence $\{x_n\}_{n\in N_+}$ taking values in $[-\infty, +\infty]$ is not bounded above (below) if there exists an $M \in \mathbb{R}$ and an $N \in N_+$ such that $x_n \leq M$ $(M \leq x_n)$ for any $n \geq N$. Suppose $\{x_n\}_{n\in N_+}$ is a sequence taking values in $[-\infty, +\infty]$, and $S^* \subset [-\infty, +\infty]$ the set of its subsequential limits.

Lemma 1.44 Let $\{x_n\}_{n \in N_+}$ be a sequence taking values in the extended real line and S^* the set of its subsequential limits. S^* contains $+\infty$ $(-\infty)$ if and only if it is not bounded above (below).

Proof) Suppose that S^* contains $+\infty$ and that $\{x_n\}_{n \in N_+}$ is bounded above. Let $M \in \mathbb{R}$ be a real number and $N \in N_+$ a natural number such that $x_n \leq M$ for any $n \geq N$. Furthermore, let $\{x_{n_k}\}_{k \in N_+}$ be a subsequence that converges to $+\infty$; then, there must exist an $k \in N_+$ such that $n_k > N$ and $x_{n_k} > M$, which contradicts the fact that $x_n \leq M$ for any $n \geq N$. Thus, $\{x_n\}_{n \in N_+}$ must not be bounded above.

Conversely, suppose that $\{x_n\}_{n \in N_+}$ is not bounded above. Then, there exists an $n_1 \in N_+$ such that $x_{n_1} > 1$, since 1 is an upper bound of $\{x_n\}_{n \in N_+}$ otherwise. Suppose that $n_1 < \cdots < n_k$ have been chosen for some $k \in N_+$. We can then choose $n_{k+1} > n_k + 1$ so that $x_{n_{k+1}} > k+1$, since $x_n \leq k+1$ for any $n \geq n_k + 1$ otherwise, which contradicts the assumption that $\{x_n\}_{n \in N_+}$ is not bounded above. The subsequence $\{x_{n_k}\}_{k \in N_+}$ constrcuted in such a manner satisfies $x_{n_k} > k$ for any $k \in N_+$, which shows us that $\{x_{n_k}\}_{k \in N_+}$ converges to $+\infty$. It follows that $+\infty \in S^*$.

The proof for when S^* contains $-\infty$ and $\{x_n\}_{n \in N_+}$ is not bounded below can be shown similarly.

Q.E.D.

By implication, if $\{x_n\}_{n \in N_+}$ is not bounded in any direction, then S^* is non-empty. On the other hand, if $\{x_n\}_{n \in N_+}$ is bounded both above and below, then $\{x_n\}_{n \geq N}$ becomes a real sequence for some $N \in N_+$ and, by the Bolzano-Weierstrass theorem, it contains a convergent subsequence. It follows that $S^* \neq \emptyset$, and thus that $S^* \neq \emptyset$ for any sequence $\{x_n\}_{n \in N_+}$ taking values in $[-\infty, +\infty]$.

We define the values $s^*, s_* \in [-\infty, +\infty]$ as follows:

$$s^* = \sup S^*, \quad s_* = \inf S^*.$$

These values are well-defined because S^* is a non-empty subset of the extended real line. We can derive the following results concerning S and S^*:

Lemma 1.45 Let $\{x_n\}_{n \in N_+}$ be a sequence taking values in $[-\infty, +\infty]$, and S^* be as defined above. Letting $S = S^* \setminus \{-\infty, +\infty\} \subset \mathbb{R}$, the following hold true:

i) S is closed in the standard topology on $\mathbb{R}$.

ii) $s^*, s_* \in S^*$, that is, there exist subsequences of $\{x_n\}_{n \in N_+}$ that converge to s^* and s_*.

Proof) i) If $S = \emptyset$, then S is trivially closed. Suppose $S \neq \emptyset$, and let s be a limit point of S. Then, there exists an $s_1 \in S$ such that $|s - s_1| < 1$. Since s_1 is the limit of a subsequence of $\{x_n\}_{n \in N_+}$, there must exist some $n_1 \in N_+$ such that $|x_{n_1} - s_1| < 1$. Assume now that $n_1 < \cdots < n_k$ have been chosen for some $k \geq 1$. Again, since s is a limit point of S, there exists a $s_{k+1} \in S$ such that $|s - s_{k+1}| < \frac{1}{k+1}$. There exists a subsequence of $\{x_n\}_{n \in N_+}$ that converges to s_{k+1}, so we can choose a natural number $n_{k+1} > n_k$ such that $|s_{k+1} - x_{n_{k+1}}| < \frac{1}{k+1}$. Letting the subsequence $\{x_{n_k}\}_{k \in N_+}$ be chosen in this manner, we can see that

$$|x_{n_k} - s| \leq |x_{n_k} - s_k| + |s_k - s| < \frac{2}{k}$$

for any $k \in N_+$, so that $\{x_{n_k}\}_{k \in N_+}$ converges to s. This makes s a real subsequential limit of S, so that $s \in S$; it follows that S is closed.

ii) We consider s^*; the claim follows symmetrically for s_*. Suppose $s^* \in \mathbb{R}$. Then, $\{x_n\}_{n \in N_+}$ must be bounded above, so that $+\infty$ is not an element of S^*, and S non-empty, so that the supremum of S^* is not $-\infty$. Consider an open interval (a, b) that contains s^*. $a < s^*$, and since $s^* = \sup S^*$, there exists an $s \in S^*$ such that $a < s \leq s^*$. This s cannot be equal to $-\infty$, and S^* does not contain $+\infty$, so $s \in S$. This shows us that $a < s \leq s^* < b$, or that $s \in S \cap (a, b)$. We have just shown that, for any open interval (a, b) on the real line containing s^*, $(a, b) \cap S \neq \emptyset$. The collection of all open intervals forms a base generating the standard topology on $\mathbb{R}$, so this result tells us that s^* is contained in the closure of S with respect to the standard topology. We saw in i) that S is closed with respect to this topology, so we ultimately have $s^* \in S \subset S^*$.

Now supose that $s^* = +\infty$, and that $\{x_n\}_{n \in N_+}$ is bounded above. This implies the existence of some $M \in \mathbb{R}$ and $N \in N_+$ such that $x_n \leq M$ for any $n \geq N$, and that S^* does not contain $+\infty$. If S^* only contains $-\infty$, then $s^* = \sup S^* = -\infty$, so S^* must also contain some real value s, or $S \neq \emptyset$. For any $s \in S$, if $s > M + 1$, then

$$|x_n - s| = s - x_n > 1$$

for any $n \geq N$, indicating that not subsequence of $\{x_n\}_{n \in N_+}$ can converge to s, which contradicts the fact that s is a subsequential limit of $\{x_n\}_{n \in N_+}$. It follows that S^* is bounded above by $M + 1$, so that

$$s^* = \sup S^* \leq M + 1.$$

This contradicts the fact that $s^* = +\infty$, so $\{x_n\}_{n \in N_+}$ must not be bounded above;

by implication S^* must contain $+\infty$, so that $s^* \in S^*$.

Finally, let $s^* = -\infty$. This indicates that S^* contains only the element $-\infty$ (since it cannot be non-empty). Thus, $s^* \in S^*$ in this case as well.

Q.E.D.

We define the limit superior and inferior of a sequence $\{x_n\}_{n \in N_+}$ taking values in the extended real line as

$$\limsup_{n\to\infty} x_n = s^*, \quad \liminf_{n\to\infty} x_n = s_*.$$

In what follows, we show that this definition is equivalent to the traditional definition using suprema and infima:

Theorem 1.46 Let $\{x_n\}_{n \in N_+}$ be a sequence taking values in $[-\infty, +\infty]$. Then,

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{k\geq n} x_k$$
$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf_{k\geq n} x_k.$$

Proof) Define S and S^* as above, and let $s^* = \limsup_{n\to\infty} x_n$. Defining $y_n = \sup_{k\geq n} x_k$ for any $n \in N_+$, the sequence $\{y_n\}_{n\in N_+}$ is a decreasing sequence, so that it converges to some $y \in [-\infty, +\infty]$; we can express y as

$$y = \lim_{n\to\infty} \sup_{k\geq n} x_k.$$

We proved above that $s^* \in S^*$, so there exists some subsequence $\{x_{n_m}\}_{m\in N_+}$ of $\{x_n\}_{n\in N_+}$ that converges to s^*. For any $m \in N_+$,

$$x_{n_m} \leq \sup_{k\geq n_m} x_k = y_{n_m},$$

so taking $m \to \infty$ on both sides yields $s^* \leq y$. If $s^* = +\infty$, it immediately follows that $s^* = y = +\infty$.

If $s^* \in \mathbb{R}$, then $\{x_n\}_{n\in N_+}$ is bounded above because S^* does not contain $+\infty$. By definition, there exist $M \in \mathbb{R}$ and $N \in N_+$ such that $x_n \leq M$ for any $n \geq N$. It follows that

$$y_n = \sup_{k\geq n} x_k \leq M$$

for any $n \geq N$ as well, which shows us that $y = \inf_{n \in N_+} y_n \leq M < +\infty$. In addition, since $-\infty < s^* \leq y$, y is real valued.

Suppose that $s^* < y$. Choosing $r \in \mathbb{Q}$ such that $s^* < r < y$, there exists an $N_1 \in N_+$ such that $N_1 \geq N$ and

$$y - y_n \leq |y_n - y| < y - r,$$

or

$$r < y_n = \sup_{k \geq n} x_k$$

for any $n \geq N_1$. Since $r < \sup_{k \geq N_1} x_k$, there exists some $n_1 \geq N_1$ such that

$$r < x_{n_1} \leq \sup_{k \geq N_1} x_k = y_{N_1} \leq M$$

Suppose that we have found $n_1 < \cdots < n_m$ for some $m \in N_+$. Similarly, because $n_m + 1 \geq N_1$, we have $r < \sup_{k \geq n_m + 1} x_k$ and we can find some $n_{m+1} \geq n_m + 1$ such that

$$r < x_{n_{m+1}} \leq \sup_{k \geq n_m + 1} x_k = y_{n_m + 1} \leq M.$$

Constructing the subsequence $\{x_{n_m}\}_{m \in N_+}$ in this manner, we can see that

$$r < x_{n_m} \leq M$$

for any $m \in N_+$. In other words, $\{x_{n_m}\}_{m \in N_+}$ is a bounded real-valued sequence, so by the Bolzano-Weierstrass theorem, it admits a convergent subsequence. Denoting this subsequential limit by $z \in S^*$, we can see that

$$s^* < r \leq z \leq M.$$

This contradicts the definition of s^* as the supremum of S^*, so we must have $s^* = y$.

Finally, if $s^* = -\infty$, then from the previous theorem, it follows that $S^* = \{-\infty\}$, and that $\{x_n\}_{n \in N_+}$ is bounded above (otherwise, $+\infty \in S^*$, a contradiction). From this we can again tell that the sequence $\{y_n\}_{n \in N_+}$ is bounded above, that is, there exist $M \in \mathbb{R}$ and $N \in N_+$ such that $y_n \leq M$ for any $n \geq N$; in particular, $y < +\infty$. Suppose $y \in \mathbb{R}$. Since $y \leq y_n$ for any $n \in N_+$, we can see that

$$-\infty < y \leq y_n \leq M$$

for any $n \geq N$, so that $y_n \in \mathbb{R}$ for any $n \geq N$.

By the definition of the supremum $y_N = \sup_{k \geq N} x_k$, we can choose an $n_1 \geq N$ such that

$$y_N - 1 < x_{n_1} \leq y_N.$$

Assume now that $n_1 < \cdots < n_m$ have been chosen for some $m \geq 1$. Then, we can choose an $n_{m+1} \geq n_m + 1$ such that

$$y_{n_m+1} - \frac{1}{m+1} < x_{n_{m+1}} \leq y_{n_m+1}.$$

The subsequence $\{x_{n_m}\}_{m \in N_+}$ constructed in this way satisfies

$$\left|x_{n_{m+1}} - y_{n_m+1}\right| < \frac{1}{m+1}$$

for any $m \in N_+$. Since $\{y_{n_m+1}\}_{m \in N_+}$ converges to y, by implication $\{x_{n_m}\}_{m \in N_+}$ converges to y as well. This means that y is a subsequential limit of $\{x_n\}_{n \in N_+}$ and is thus contained in S^*. But this contradicts the fact that S^* consists only of the element $-\infty$, so it must be the case that $s^* = y = -\infty$.

The claim for limit inferior follows by observing that

$$\begin{aligned}
\liminf_{n \to \infty} x_n &= -\limsup_{n \to \infty}(-x_n) \\
&= -\lim_{n \to \infty}\left(\sup_{k \geq n}(-x_k)\right) \\
&= \lim_{n \to \infty}\left(-\sup_{k \geq n}(-x_k)\right) = \lim_{n \to \infty} \inf_{k \geq n} x_k.
\end{aligned}$$

Q.E.D.

The main use of limit superior and inferior in analysis arises when proving the existence of the limit of a sequence taking values in the extended real line. We show now that a sequence in $[-\infty, +\infty]$ converges to the limit superior (inferior) if and only if the two values are equal. This furnishes us with a convenient characterization of convergence of sequences taking values in the extended real line that still holds for real valued sequences.

Theorem 1.47 Let $\{x_n\}_{n \in N_+}$ be a sequence taking values in $[-\infty, +\infty]$. For any $-\infty \leq x \leq +\infty$, $\{x_n\}_{n \in N_+}$ conveges to x if and only if

$$\limsup_{n \to \infty} x_n = \liminf_{n \to \infty} x_n = x.$$

Proof) **Sufficiency**

Note that $S^* = \{x\}$ if

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n = x.$$

We consider three cases. First, suppose that $x \in \mathbb{R}$. In this case, $\{x_n\}_{n\in N_+}$ is both bounded above (otherwise $+\infty \in S^*$) and below (otherwise $-\infty \in S^*$). We can tell from this that there exists some $N \in N_+$ such that $\{x_n\}_{n\geq N}$ is real-valued, and can be considered a sequence in the metric space $(\mathbb{R}, d)$, where d is the euclidean metric on $\mathbb{R}$.

Consider any subsequence of $\{x_n\}_{n\geq N}$; it is a bounded real sequence, so by Bolzano-Weierestrass, this subsequence has at least one convergent subsequence, and its limit is x because it is also a subsequence of $\{x_n\}_{n\in N_+}$ and x is the only subsequential limit. Therefore, any subsequence of the real sequence $\{x_n\}_{n\geq N}$ has a further subsequence that converges to x, and by the selection theorem, $\{x_n\}_{n\geq N}$ itself converges to x, which is the same as saying that $\{x_n\}_{n\in N_+}$ itself converges to x.

On the other hand, suppose that $x = +\infty$. In this case,

$$\liminf_{n\to\infty} x_n = \sup_{n\in N_+} \inf_{k\geq n} x_k = +\infty.$$

For any $M \in \mathbb{R}$, this means that there exists some $N \in N_+$ such that

$$x_k \geq \inf_{k\geq N} x_k > M$$

for any $k \geq N$. This holds for any $M \in \mathbb{R}$, so by definition $\{x_n\}_{n\in N_+}$ converges to $+\infty$. We can show similarly that the same result holds when $x = -\infty$ by making use of the limit superior.

Necessity

Suppose that $\{x_n\}_{n\in N_+}$ converges to x. This indicates that every subsequence of $\{x_n\}_{n\in N_+}$ converges to x (checking this is very easy for the three cases $x \in \mathbb{R}$, $x = +\infty$ and $x = -\infty$). Thus, $S^* = \{x\}$ and

$$\limsup_{n\to\infty} x_n = \sup S^* = x = \inf S^* = \liminf_{n\to\infty} x_n.$$

Q.E.D.

1.17.2 Completeness of Euclidean Spaces

The limits superior and inferior can be used to easily prove that the euclidean spaces $\mathbb{R}^n$ are complete metric spaces, that is, any Cauchy sequence in $\mathbb{R}^n$ is also convergent. We first prove this for the real line $\mathbb{R}$ equipped with the euclidean metric.

Theorem 1.48 (Completeness of the Real Line)
The real line $\mathbb{R}$ is a complete metric space under the euclidean metric.

Proof) Let $\{x_n\}_{n\in N_+}$ be a Cauchy sequence in $\mathbb{R}$. It is easy to see that $\{x_n\}_{n\in N_+}$ is bounded; by definition, there exists an $N \in N_+$ such that $|x_n - x_m| < 1$ for any $n, m \geq N$, so

$$|x_n| \leq |x_n - x_N| + |x_N| < 1 + |x_N|$$

for any $n \geq N$. Defining

$$M = \max(1 + |x_N|, |x_1|, \cdots, |x_{N-1}|) < +\infty,$$

we can now see that

$$|x_n| \leq M$$

for any $n \in N_+$, implying that $\{x_n\}_{n\in N_+}$ is a bounded sequence. Therefore, its limit superior and inferior are both real numbers:

$$s^* = \limsup_{n\to\infty} x_n, \quad s_* = \liminf_{n\to\infty} x_n \in \mathbb{R}.$$

We need only show that $s^* = s_*$ to show, in light fo the preceding theorem, that $\{x_n\}_{n\in N_+}$ converges to this common (real) value.

For any $\epsilon > 0$, by the Cauchy property of $\{x_n\}_{n\in N_+}$ and the fact that

$$s^* = \inf_{n\in N_+} \sup_{k\geq n} x_k, \quad s_* = \sup_{n\in N_+} \inf_{k\geq n} x_k,$$

there exists an $N \in N_+$ such that

$$|x_n - x_m| < \epsilon$$

and

$$\left|\sup_{k\geq n} x_k - s^*\right| < \epsilon, \quad \text{and} \quad \left|\inf_{k\geq n} x_k - s_*\right| < \epsilon$$

for any $n, m \geq N$.

Fix $n \geq N$. Then, for any $k \geq n$, since $k, n \geq N$, we have

$$x_k - x_n \leq |x_k - x_n| < \epsilon,$$

so that

$$\sup_{k \geq n} (x_k - x_n) = \sup_{k \geq n} x_k - x_n \leq \epsilon.$$

Similarly, since

$$x_n - x_k \leq |x_k - x_n| < \epsilon$$

for any $k \geq N$, it follows that

$$\sup_{k \geq n} (x_n - x_k) = x_n - \inf_{k \geq n} x_k \leq \epsilon$$

as well. Adding together the two inequalities shows us that

$$0 \leq \sup_{k \geq n} x_k - \inf_{k \geq n} x_k \leq 2\epsilon.$$

In addition, we can also see that, since $n \geq N$,

$$\left| \sup_{k \geq n} x_k - s^* \right| < \epsilon, \quad \text{and} \quad \left| \inf_{k \geq n} x_k - s_* \right| < \epsilon.$$

It follows that

$$|s^* - s_*| \leq \left| \sup_{k \geq n} x_k - s^* \right| + \left| \sup_{k \geq n} x_k - \inf_{k \geq n} x_k \right| + \left| \inf_{k \geq n} x_k - s_* \right| \leq 4\epsilon.$$

This holds for any $\epsilon > 0$, so we can conclude that $s^* = s_* = s \in \mathbb{R}$, and therefore that $\{x_n\}_{n \in N_+}$ converges to s.

Q.E.D.

Since the convergence of sequences in $\mathbb{R}^n$ is equivalent to the convergence of each of its coordinates in $\mathbb{R}$, the completeness of arbitrary euclidean spaces follows as a simple corollary of the preceding result:

Corollary to Theorem 1.48 (Completeness of Euclidean Spaces)
The euclean n-space $\mathbb{R}^n$ is a complete metric space under the euclidean metric.

Proof) Let $\{x_k\}_{k \in N_+}$ be a Cauchy sequence in $\mathbb{R}^n$, and denote each x_k as $x_k = (x_{1,k}, \cdots, x_{n,k}) \in \mathbb{R}^n$. Note that each sequence $\{x_{i,k}\}_{k \in N_+}$ in $\mathbb{R}$ is a Cauchy sequence; this is because, for any $k, m \in N_+$,

$$|x_{i,k} - x_{i,m}| \leq \left(\sum_{j=1}^{n} |x_{j,k} - x_{j,m}|^2 \right)^{\frac{1}{2}} = |x_k - x_m|$$

for any $1 \leq i \leq n$. By the completeness of $\mathbb{R}$, for any $1 \leq i \leq n$ there exists an $x_i^* \in \mathbb{R}$ such that $x_{i,k} \to x_i^*$ as $k \to \infty$. Defining $x^* = (x_1^*, \cdots, x_n^*) \in \mathbb{R}^n$, note that

$$|x_k - x^*| = \left(\sum_{i=1}^{n} |x_{i,k} - x_i^*|^2 \right)^{\frac{1}{2}},$$

so that taking $k \to \infty$ on both sides shows us $\{x_k\}_{k \in N_+}$ converges to x^*.

Q.E.D.

The completeness of euclidean spaces also allows us to recover the Heine-Borel theorem as a special case of the Borel-Lebesgue lemma. To understand this theorem, we need to introduce the concept of ϵ-nets of sets. Given a metric space (E, d) and a set $A \subset E$, we say that the set $B \subset E$ is an ϵ-net of A for some $\epsilon > 0$ if

$$A \subset \bigcup_{x \in B} B_d(x, \epsilon),$$

that is, we can approximate A with accuracy ϵ using the points in B. Note that we need not require the points in B to be contained in A; this is convenient, because it shows that any ϵ-net of a set A is also an ϵ-net of any subset of A. To see this, let A and A' be sets such that $A' \subset A$, and let $B \subset E$ be an ϵ-net of A. In this case,

$$A' \subset A \subset \bigcup_{x \in B} B_d(x, \epsilon),$$

so that B is also an ϵ-net of A'.

We say that B is a finite ϵ-net of A if B is a finite set. If a set A has a finite ϵ-net for any $\epsilon > 0$, then we say that A is totally bounded.

Theorem 1.49 (Borel-Lebesgue Lemma)
Let (E, d) be a metric space. If a subset A of E is compact, then it is closed and totally bounded. If, in addition, (E, d) is complete, then the converse holds as well.

Proof) **Necessity**

Suppose that A is a compact set. Since metric spaces are Hausdorff spaces, A is a closed set. Suppose that there exists some $\epsilon > 0$ such that A has no finite ϵ-net. Choose any $x_1 \in A$. Then, assuming that we have chosen $x_1, \cdots, x_k \in A$ for some $k \geq 1$, we can choose $x_{k+1} \in A$ as

$$x_{k+1} \in A \setminus \left(\bigcup_{i=1}^{k} B_d(x_i, \epsilon) \right);$$

we can find such an y_{k+1} because otherwise, A would have $\{x_1, \cdots, x_k\}$ as a finite ϵ-net. Having constructed $\{x_k\}_{k \in N_+} \subset A$ as such, note that, for any $n, m \in N_+$, assuming $n < m$ without loss of generality,

$$d(x_m, x_n) \geq \epsilon,$$

since otherwise, $x_m \in B_d(x_n, \epsilon)$, a contradiction. On the other hand, since $\{x_k\}_{k \in N_+}$ is a sequence in the compact (and thus sequentially compact) set A, it has a subsequence that converges to some point in $x \in A$. This subsequence, being convergent, must also be Cauchy, but we just proved above that no subsequence of $\{x_k\}_{k \in N_+}$ can be Cauchy (the distance between any two elements in the sequence is bounded below by ϵ). This results in a contradiction, and thus it must be the case that A has a finite ϵ-net for any $\epsilon > 0$.

Sufficiency

Now suppose that A is closed and totally bounded. We use the sequential compactness characterization to prove that A is compact. Choose any sequence $\{x_n\}_{n \in N_+}$ in A. Then, there must exist an $y_1 \in E$ such that the open ball $B_d(x, 1)$ contains infinitely many elements of $\{x_n\}_{n \in N_+}$; otherwise, A admits no finite 1-net, which contradicts our initial assumption.

Assume that we have chosen $y_1, \cdots, y_k \in E$ for some $k \geq 1$ such that $\bigcap_{i=1}^{k} B_d(y_i, 1/i)$ contains infinitely many elements of $\{x_n\}_{n \in N_+}$. Then, we choose y_{k+1} so that the intersection

$$\bigcap_{i=1}^{k+1} B_d(y_i, 1/i)$$

also contains in infinitely many elements in $\{x_n\}_{n \in N_+}$. To see why this choice is possible, note that $\bigcap_{i=1}^{k} B_d(y_i, 1/i)$ contains infinitely many elements of $\{x_n\}_{n \in N_+}$; thus there exists an $N \in N_+$ such that $\{x_n\}_{n \geq N} \subset \bigcap_{i=1}^{k} B_d(y_i, 1/i)$. Then, $\{x_n\}_{n \geq N}$, being a subset of A, is also totally bounded and therefore it has a finite $\frac{1}{k+1}$-net by assumption. In other words, at least one of the $\frac{1}{k+1}$-balls that covers $\{x_n\}_{n \geq N}$ must contain infinitely many elements of $\{x_n\}_{n \geq N}$.

Now we choose $n_1 \in N_+$ so that $x_{n_1} \in B_d(y_1, 1)$; this is possible because $B_d(y_1, 1)$ contains elements of $\{x_n\}_{n \in N_+}$. Assume that we have chosen $n_1 < \cdots < n_k$ for some $k \geq 1$;

then, choose $n_{k+1} > n_k$ so that

$$x_{n_{k+1}} \in \bigcap_{i=1}^{k+1} B_d(y_i, 1/i).$$

Again, we can choose to make the index n_{k+1} larger than n_k because the intersection on-the right hand side contains infinitely many elements of $\{x_n\}_{n\in N_+}$. Having constructed the subsequence $\{x_{n_k}\}_{k\in N_+}$ as above, note that, for any $k, l \in N_+$ such that $l > k$,

$$x_{n_k}, x_{n_l} \in \bigcap_{i=1}^{k} B_d(y_i, 1/i) \subset B_d(y_k, 1/k),$$

and as such

$$d(x_{n_k}, x_{n_l}) \leq d(x_{n_k}, y_k) + d(x_{n_l}, y_k) < \frac{2}{k}.$$

Thus,

$$\lim_{k,l\to\infty} d(x_{n_k}, x_{n_l}) = 0,$$

making $\{x_{n_k}\}_{k\in N_+}$ a sequence that is Cauchy in d. By the completeness of (E, d), it follows that this subsequence converges to some point $x^* \in E$, and since $\{x_{n_k}\}_{k\in N_+}$ is a sequence in A, a closed set, $x^* \in A$. We have shown that any sequence in A has a subsequence that converges to some point in A; by definition, A is sequentially compact, and since we are working with metric spaces, A is compact as well.

Q.E.D.

The Heine-Borel theorem now follows easily:

Theorem 1.50 (Heine-Borel Theorem)
A subset A of the euclidean n-space $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof) Let d denote the euclidean metric on $\mathbb{R}^n$. It suffices, in light of the Borel-Lebesgue theorem, to show that A is totally bounded if and only if it is bounded. Suppose that A has a finite ϵ-net for any $\epsilon > 0$. Then, there exists a finite set $\{x_1, \cdots, x_k\} \subset \mathbb{R}^n$ such that

$$A \subset \bigcup_{i=1}^{k} B_d(x_i, 1).$$

Defining $M = \max(|x_1| + 1, \cdots, |x_k| + 1) < +\infty$, note that, for any $x \in A$, letting $x \in$

$B_d(x_i,1)$ for some $1 \leq i \leq k$,

$$|x| \leq |x - x_i| + |x_i| < 1 + |x_i| \leq M.$$

As such, A is a bounded set.

Conversely, suppose that A is a bounded set. Then, there exists an $0 < M < +\infty$ such that $|x| \leq M$ for any $x \in A$, so that A is contained in the n-cell $[-M, M]^n$. For any $\epsilon > 0$, choose $N \in N_+$ so that $\frac{1}{N} < \frac{\epsilon}{\sqrt{n}}$. Then, let E be the set

$$E = \left\{ x \in \mathbb{R}^n \mid \forall 1 \leq i \leq k,\ x_i = -M + \frac{1}{N} \cdot j \text{ for some } 0 \leq j \leq 2MN \right\}.$$

In other words, we are subdividing the n-cell $[-M, M]^n$ into identically sized cubes whose edge length is $\frac{1}{N}$. Clearly, E is a finite subset of $\mathbb{R}^n$. We can then see that

$$A \subset \bigcup_{y \in E} B_d(y, \epsilon);$$

to see this, choose any $x \in A$. Then, there exist $j_1, \cdots, j_n \in \{0, \cdots, 2MN\}$ such that

$$x \in \prod_{i=1}^{n} \left[-M + \frac{1}{N} \cdot j_i,\ -M + \frac{1}{N} \cdot (j_i + 1) \right].$$

Defining

$$y = \left(-M + \frac{1}{N} \cdot j_1, \cdots, -M + \frac{1}{N} \cdot j_n \right) \in E,$$

for any $1 \leq i \leq n$ we have

$$|x_i - y_i| \leq \frac{1}{N} < \frac{\epsilon}{\sqrt{n}},$$

so that

$$|x - y| = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{\frac{1}{2}} < \epsilon$$

and $x \in B_d(y, \epsilon)$. We can construct such a cover of A for any $\epsilon > 0$, so by definition A is totally bounded.

Since $(\mathbb{R}^n, d)$ is a complete metric space, A is compact if and only if A is closed and totally bounded by the Borel-Lebesgue lemma. Since the second condition and the boundedness of A are equivalent for euclidean spaces, it follows that A is compact if and only if A is closed and bounded.

Q.E.D.

Chapter 2

Measure Spaces and Measurable Functions

2.1 Measurable Spaces and Borel Spaces

2.1.1 Measurable Spaces and Generating Sets

Let E be any set, and $\mathcal{A}$ a collection of subsets of E. $\mathcal{A}$ is said to be an algebra on E if it satisfies the following properties:

i) $E \in \mathcal{A}$; $\mathcal{A}$ contains the entire set

ii) For any $A \in \mathcal{A}$, $A^c = E \setminus A \in \mathcal{A}$; $\mathcal{A}$ is closed under complements

iii) For any finite collection $\{A_1, \cdots, A_n\} \subset \mathcal{A}$, the union $A = \bigcup_{i=1}^n A_i \in \mathcal{A}$; $\mathcal{A}$ is closed under finite unions.

It follows that, $\emptyset = E^c \in \mathcal{A}$ because $\mathcal{A}$ is closed under complementation, and that, for a finite collection of sets $\{A_1, \cdots, A_n\} \subset \mathcal{A}$,

$$\bigcap_{i=1}^n A_i = \left(\bigcup_{i=1}^n A_i^c \right)^c \in \mathcal{A}$$

because $\mathcal{A}$ is closed under complementation and finite unions. Therefore, $\mathcal{A}$ is also closed under finite intersections.

The collection $\mathcal{E}$ of subsets of E is said to be a σ-algebra on E if it satisfies the following properties:

i) $E \in \mathcal{E}$; $\mathcal{E}$ contains the entire set

ii) For any $A \in \mathcal{E}$, $A^c = E \setminus A \in \mathcal{E}$; $\mathcal{E}$ is closed under complements

iii) For any countable collection $\{A_n\}_{n \in N_+} \subset \mathcal{E}$, the union $A = \bigcup_n A_n \in \mathcal{E}$; $\mathcal{E}$ is closed under countable unions

Clearly, $\mathcal{E}$ is simply an algebra on E that is closed under countable unions as well as finite unions.
The elements of $\mathcal{E}$ are called measurable sets, and the pair $(E,\mathcal{E})$ is called a measurable space.

Let $(E,\mathcal{E})$ be a measurable space. The following directly follow from the definition of a σ-algebra:

- $\emptyset \in \mathcal{E}$ because $\emptyset = E \setminus E$, $E \in \mathcal{E}$ and $\mathcal{E}$ is closed under complements
- For any $A_1, \cdots, A_n \in \mathcal{E}$, defining $A_i = \emptyset$ for any $i \geq n+1$,

  $$\bigcup_{i=1}^{n} A_i = \bigcup_m A_m \in \mathcal{E}$$

  because each $A_i \in \mathcal{E}$ and $\mathcal{E}$ is closed under countable unions.
- For any countable collection $\{A_n\}_{n\in N_+}$, because $A_n^c \in \mathcal{E}$ for any $n \in N_+$,

  $$\bigcap_n A_n = \left(\bigcup_n A_n^c\right)^c \in \mathcal{E}.$$

  By the same logic as the point above, $\mathcal{E}$ is also closed under finite intersections.
- For any $A, B \in \mathcal{E}$,

  $$A \setminus B = A \cap B^c \in \mathcal{E}$$

  because $\mathcal{E}$ is closed under finite intersections and complements.

Like with topologies, we often characterize σ-algebras using some collection of subsets contained in $\mathcal{E}$. The exact sense in which the characterization is possible is explained below.

Lemma 2.1 The intersection of σ-algebras is also a σ-algebra.

Proof) Let $\{\mathcal{E}_\alpha\}$ be an arbitrary collection of σ-algebras on E, and define $\mathcal{E} = \bigcap_\alpha \mathcal{E}_\alpha$.

We now show that $\mathcal{E}$ satisfies the three conditions for a σ-algebra:

i) The entire set E is contained in each $\mathcal{E}_\alpha$, so it is contained in their intersection $\mathcal{E}$ as well.

ii) For any $A \in \mathcal{E}$, since $A \in \mathcal{E}_\alpha$ for all α, the complement $A^c = E \setminus A \in \mathcal{E}_\alpha$ as well. As such, $A^c \in \mathcal{E}$.

iii) For any countable collection $\{A_n\}_{n\in N_+} \in \mathcal{E}$, since $\{A_n\}_{n\in N_+} \in \mathcal{E}_\alpha$ for all α, the union $A = \bigcup_n A_n \in \mathcal{E}_\alpha$ as well. As such, $A \in \mathcal{E}$.

Therefore, by definition, $\mathcal{E}$ is a σ-algebra on E.
Q,E.D.

For any collection $\mathcal{F}$ of subsets of E, the σ-algebra generated by $\mathcal{F}$ is defined as the intersection of all σ-algebras on E that contain $\mathcal{F}$, and denoted $\sigma\mathcal{F}$.
It is in this sense that $\sigma\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{F}$.

2.1.2 Borel Spaces

Let (E,τ) be a topological space. Then, the σ-algebra generated by the collection τ of subsets of E is called the Borel σ-algebra on (E,τ), and denoted $\mathcal{B}(E,\tau)$. The elements of $\mathcal{B}(E,\tau)$ are referred to as Borel sets.
We will often encounter the Borel σ-algebras $\mathcal{B}(\mathbb{R},\tau_{\mathbb{R}})$, $\mathcal{B}([-\infty,+\infty],\tau_{[-\infty,+\infty]})$ and $\mathcal{B}(\mathbb{R}^n,\tau_n^e)$ on euclidean spaces. For the sake of notational convenience, we now denote them as $\mathcal{B}(\mathbb{R})$, $\mathcal{B}([-\infty,+\infty])$ and $\mathcal{B}(\mathbb{R}^n)$.

The following relates the base of a second countable topological space and the Borel σ-algebra it generates; it is an example of the usefulness of second countability.

Lemma 2.2 Let (E,τ) be a second countable topological space and $\mathbb{B}$ a countable base of E that generates τ. Then, $\mathcal{B}(E,\tau)=\sigma\mathbb{B}$, that is, the base $\mathbb{B}$ generates the Borel σ-algebra $\mathcal{B}(E,\tau)$.

Proof) Let $\mathcal{E}$ be the σ-algebra generated by $\mathbb{B}$. We must show that $\mathcal{E}=\mathcal{B}(E,\tau)$.
Since τ is contained in the σ-algebra $\mathcal{B}(E,\tau)$ and $\mathbb{B}$ is contained in τ, $\mathcal{B}(E,\tau)$ is a σ-algebra contaning $\mathbb{B}$. $\mathcal{E}$ is the smallest σ-algebra containing $\mathbb{B}$ by definition, so $\mathcal{E}\subset\mathcal{B}(E,\tau)$.

To see the reverse inclusion, choose any $A\in\tau$. Because $\mathbb{B}$ is a base that generates τ, by theorem 1.3, $A=\bigcup_i B_i$ for some colleciton $\{B_i\}\subset\mathbb{B}$; but $\mathbb{B}$ is countable, so that A must be the countable union of sets in $\mathbb{B}$. Sets in $\mathbb{B}$ are also contained in $\mathcal{E}$, so A is the countable union of sets in $\mathcal{E}$. Finally, σ-algebras are closed under countable unions, which means that $A\in\mathcal{E}$.
This holds for any $A\in\tau$, so $\tau\subset\mathcal{E}$. By definition, $\mathcal{B}(E,\tau)$ is the smallest σ-algebra containing τ, so it must be the case that $\mathcal{B}(E,\tau)\subset\mathcal{E}$.
Therefore, $\mathcal{B}(E,\tau)=\mathcal{E}=\sigma\mathbb{B}$.
Q.E.D.

It follows from the above result and theorem 1.5 that the collection of all open intervals with rational endpoints generates $\mathcal{B}(\mathbb{R})$.

2.1.3 The $\pi-\lambda$ Theorem and Monotone Classes of Sets

Here we introduce special concepts related to measurable spaces that are frequently used in probability theory.
Let E be an arbitrary set. We say that a collection $\mathcal{E}_0$ of subsets of E is a π-system if it is closed under finite intersections, that is, if:

$$\text{For any } A, B \in \mathcal{E}_0,\ A \cap B \in \mathcal{E}_0.$$

On the other hand, we say that a collection $\mathcal{M}$ of subsets of E is a λ-system if:

i) $E \in \mathcal{M}$

ii) For any $A, B \in \mathcal{M}$ such that $A \subset B$, $B \setminus A \in \mathcal{M}$

iii) For any sequence $\{A_n\}_{n \in N_+} \in \mathcal{M}$ such that $A_n \subset A_{n+1}$ for any $n \in N_+$, the set $A = \bigcup_n A_n \in \mathcal{M}$.

Note that any σ-algebra is both a π-system and a λ-system.
We can prove the converse of this statement as well, namely that, if a collection of subsets of E is both a π-system and a λ-system, then it is a σ-algebra on E.

Lemma 2.3 Let E be a set and $\mathcal{E}$ a collection of subsets of E. If $\mathcal{E}$ is both a π-system and a λ-system, then it is a σ-algebra on E.

Proof) We will check the criteria for a σ-algebra one by one:

i) $E \in \mathcal{E}$ because $\mathcal{E}$ is a λ-system.

ii) Let $A \in \mathcal{E}$. Since $E \in \mathcal{E}$ and $A \subset E$, by the second property of a λ-system we have $A^c = E \setminus A \in \mathcal{E}$.

iii) Let $\{A_n\}_{n \in N_+}$ be an arbitrary collection of sets in $\mathcal{E}$. Define $\{B_n\}_{n \in N_+}$ as $B_1 = A_1$ and

$$B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i \right)$$

for any $n \geq 2$. Then, $\{B_n\}_{n \in N_+}$ is an increasing sequence of sets, where $B_1 \in \mathcal{E}$ trivially. For any $n \geq 2$, because

$$B_n = A_n \cap \left(\bigcup_{i=1}^{n-1} A_i \right)^c = A_n \cap \left(\bigcap_{i=1}^{n-1} A_i^c \right),$$

$B_n \in \mathcal{E}$ because it is closed under complements (ii) above) and finite intersections ($\mathcal{E}$ is a π-system).

Therefore, $\{B_n\}_{n\in N_+} \subset \mathcal{E}$, and by the third property of λ-systems, $\bigcup_n A_n = \bigcup_n B_n \in \mathcal{E}$.

By definition, $\mathcal{E}$ is a σ-algebra on E.
Q.E.D.

The next result, the $\pi-\lambda$ Theorem, furnishes us with a convenient way to verify that a σ-algebra is part of some collection of subsets. First, note that the intersection of π- and λ-systems are also π- and λ-systems, much like how the intersection of topologies and σ-algebras are also topologies and σ-algebras.

Theorem 2.4 (The $\pi-\lambda$ Theorem)
Let E be a set and $\mathcal{E}_0$ a π-system on E. If $\mathcal{M}$ is a λ-system on E that contains $\mathcal{E}_0$, then $\mathcal{M}$ also contains the σ-algebra $\mathcal{E} = \sigma\mathcal{E}_0$ generated by $\mathcal{E}_0$.

Proof) Let $\mathcal{M}_0$ be the smallest λ-system containing $\mathcal{E}_0$, that is, the intersection of all λ-systems containing $\mathcal{E}_0$. We will show that $\mathcal{M}_0$ is precisely the σ-algebra $\mathcal{E}$ generated by $\mathcal{E}_0$, at which point the proof will be complete.

Suppose that $\mathcal{M}_0$ is a π-system. Then, the preceding lemma tells us that $\mathcal{M}_0$ is a σ-algebra on E. $\mathcal{M}_0$ contains $\mathcal{E}_0$, so that $\mathcal{E} \subset \mathcal{M}_0$ Conversely, $\mathcal{E}$ is a λ-system containing $\mathcal{E}_0$, so $\mathcal{M}_0 \subset \mathcal{E}$ by the definition of $\mathcal{M}$, which shows that $\mathcal{M}_0 = \mathcal{E} = \sigma\mathcal{E}_0$.

Therefore, the proof will be complete if we just show that $\mathcal{M}_0$ is a π-system. To this end, choose any $A \in \mathcal{E}_0$, and define

$$\mathcal{M}_A = \{B \subset E \mid A \cap B \in \mathcal{M}_0\}.$$

We can show that $\mathcal{M}_A$ is a λ-system on E:

i) $E \in \mathcal{M}_A$ because $E \cap A = A \in \mathcal{M}_0$.

ii) For any $B_1, B_2 \in \mathcal{M}_A$ such that $B_1 \subset B_2$,

$$A \cap (B_2 \setminus B_1) = (A \cap B_2) \setminus (A \cap B_1);$$

because $A \cap B_2, A \cap B_1 \in \mathcal{M}_0$, $A \cap B_1 \subset A \cap B_2$ and $\mathcal{M}_0$ is a λ-system, their difference is also in $\mathcal{M}_0$. Therefore, $B_2 \setminus B_1 \in \mathcal{M}_A$.

iii) For any increasing sequence $\{B_n\}_{n\in N_+} \in \mathcal{M}_A$, define $B = \bigcup_n B_n$. Then,

$$A \cap B = A \cap \left(\bigcup_n B_n\right) = \bigcup_n (A \cap B_n).$$

Since each $A \cap B_n \in \mathcal{M}_0$, $A \cap B_n \subset A \cap B_{n+1}$ and $\mathcal{M}_0$ is a λ-system, $A \cap B \in \mathcal{M}_0$ as well, and $B \in \mathcal{M}_A$.

$\mathcal{M}_A$ is thus a λ-system on E that contains the π-system $\mathcal{E}_0$. By the definition of $\mathcal{M}_0$, $\mathcal{M}_0 \subset \mathcal{M}_A$, so that $A \cap B \in \mathcal{M}_0$ for any $B \in \mathcal{M}_0$. Thsi holds for any $A \in \mathcal{E}_0$, so $A \cap B \in \mathcal{M}_0$ for any $A \in \mathcal{E}_0$ and $B \in \mathcal{M}_0$.

Now define the collection

$$\mathcal{M}_1 = \{A \subset E \mid A \cap B \in \mathcal{M}_0 \text{ for any } B \in \mathcal{M}_0\}.$$

We can show that this is also a λ-system on E:

i) $E \cap B = B \in \mathcal{M}_0$ for any $B \in \mathcal{M}_0$, so $E \in \mathcal{M}_1$.

ii) For any $A_1, A_2 \in \mathcal{M}_1$ such that $A_1 \subset A_2$, for any $B \in \mathcal{M}_0$ we have

$$B \cap (A_1 \setminus A_2) = (B \cap A_1) \setminus (B \cap A_2).$$

Because $A_1 \cap B, A_2 \cap B \in \mathcal{M}_0$, $A_1 \cap B \subset A_2 \cap B$ and $\mathcal{M}_0$ is a λ-system, their difference is also in $\mathcal{M}_0$. Therefore, $A_2 \setminus A_1 \in \mathcal{M}_1$.

iii) For any increasing sequence $\{A_n\}_{n \in N_+} \in \mathcal{M}_1$, define $A = \bigcup_n A_n$ and choose some $B \in \mathcal{M}_0$. Then,

$$A \cap B = \left(\bigcup_n A_n\right) \cap B = \bigcup_n (A_n \cap B).$$

Since each $A_n \cap B \in \mathcal{M}_0$, $A_n \cap B \subset A_{n+1} \cap B$ and $\mathcal{M}_0$ is a λ-system, $A \cap B \in \mathcal{M}_0$ as well, and $A \in \mathcal{M}_1$.

$\mathcal{M}_1$ is thus a λ-system containing $\mathcal{E}_0$; therefore, as before, we can see that $\mathcal{M}_0 \subset \mathcal{M}_1$, or that $A \cap B \in \mathcal{M}_0$ for any $A, B \in \mathcal{M}_0$. It follows that $\mathcal{M}_0$ is a π-system, and the proof is complete.
Q.E.D.

Finally, we introduce the mathematical object known as a monotone class. The collection $\mathcal{M}$ of subsets of E is said to be a monotone class of sets if:

i) For any sequence $\{A_n\}_{n\in N_+}$ in $\mathcal{M}$ such that $A_n \subset A_{n+1}$ for any $n \in N_+$,

$$\bigcup_n A_n \subset \mathcal{M}.$$

ii) For any sequence $\{B_n\}_{n\in N_+}$ in $\mathcal{M}$ such that $B_{n+1} \subset B_n$ for any $n \in N_+$,

$$\bigcap_n B_n \subset \mathcal{M}.$$

In other words, $\mathcal{M}$ is a monotone class of sets if it is closed under increasing countable unions and decreasing countable intersections. Clearly, σ-algebras are monotone classes, since they are closed under arbitrary countable unions and intersections.

As with σ-algebras and λ-systems, the arbitrary intersection of montone classes is also a monotone class (this can be shown very easily, through the same process as in lemma 2.1). As such, for any collection $\mathcal{A}$ of subsets of E, we can define the smallest monotone class containing $\mathcal{A}$ as the intersection of all montone classes containing $\mathcal{A}$.

Montone classes of sets are useful because of the following result, which is called the monotone class theorem; it can be shown as a corollary to the $\pi - \lambda$ theorem.

Corollary to Theorem 2.4 (The Montone Class Theorem)

Let E be a set and $\mathcal{A}$ an algebra on E. If $\mathcal{M}$ is the smallest monotone class containing $\mathcal{A}$, then $\mathcal{M}$ is precisely the σ-algebra generated by $\mathcal{A}$.

Proof) The σ-algebra generated by $\mathcal{A}$ is a monotone class containing $\mathcal{A}$ and therefore contains $\mathcal{M}$.

To show the reverse inclusion, we first note that $\mathcal{M}$ possess the following properties:

- Because $\mathcal{A} \subset \mathcal{M}$ and $E \in \mathcal{A}$ by the definition of an algebra on E, it follows that $E \in \mathcal{M}$ as well.
- For any increasing sequence of sets $\{A_n\}_{n\in N_+} \subset \mathcal{M}$, letting $A = \bigcup_n A_n$, $A \in \mathcal{M}$ by definition of a montone class of sets.

Therefore, we need only show that $B \setminus A \in \mathcal{M}$ for any $A, B \in \mathcal{M}$ such that $A \subset B$ for $\mathcal{M}$ to be a λ-system on E.

To this end, let $A \in \mathcal{A}$ and define

$$\mathcal{M}_1 = \{B \subset E \mid B \setminus A \in \mathcal{M}\}.$$

For any increasing sequence $\{B_n\}_{n\in N_+} \subset \mathcal{M}$, letting $B = \bigcup_n B_n$, because $B_n \setminus A \in \mathcal{M}$ for any $n \in N_+$ and $\{B_n \setminus A\}_{n\in N_+}$ is an increasing sequence of sets in $\mathcal{M}$, by the first property of montone classes we have

$$B \setminus A = \bigcup_n (B_n \setminus A) \in \mathcal{M}.$$

Likewise, for any decreasing sequence $\{B_n\}_{n\in N_+} \subset \mathcal{M}$, letting $B = \bigcap_n B_n$, because $B_n \setminus A \in \mathcal{M}$ for any $n \in N_+$ and $\{B_n \setminus A\}_{n\in N_+}$ is an decreasing sequence of sets in $\mathcal{M}$, by the second property of montone classes we have

$$B \setminus A = \bigcap_n (B_n \setminus A) \in \mathcal{M}.$$

By definition, $\mathcal{M}_1$ is a monotone class of sets on E, and because

$$B \setminus A = B \cap A^c \in \mathcal{A} \subset \mathcal{M}$$

for any $B \in \mathcal{A}$, $\mathcal{M}_1$ is a monotone class containing the algebra $\mathcal{A}$. By the definition of $\mathcal{M}$ as the smallest monotone class containing $\mathcal{A}$, we have $\mathcal{M} \subset \mathcal{M}_1$, so that $B \setminus A \in \mathcal{M}$ for any $B \in \mathcal{M}$.

Now define

$$\mathcal{M}_2 = \{A \subset E \mid B \setminus A \in \mathcal{M} \text{ for any } B \in \mathcal{M}\}.$$

Let $\{A_n\}_{n\in N_+}$ be an increasing sequence of sets in $\mathcal{M}_2$ with limit $A = \bigcup_n A_n$. Choose any $B \in \mathcal{M}$. Then, because $B \setminus A_n \in \mathcal{M}$ for any $n \in N_+$, $\{B \setminus A_n\}_{n\in N_+}$ is a decreasing sequence of sets in $\mathcal{M}$ with limit $B \setminus A$. It follows by the definition of a montone class that $B \setminus A \in \mathcal{M}$. Therefore, $A \in \mathcal{M}_2$.
Likewise, for any decreasing sequence of sets $\{A_n\}_{n\in N_+}$ in $\mathcal{M}_2$ with limit A, for any $B \in \mathcal{M}$, $\{B \setminus A_n\}_{n\in N_+}$ is an increasing sequence of sets in $\mathcal{M}$, so that its limit $B \setminus A$ is also in $\mathcal{M}$. As such, $A \in \mathcal{M}_2$, and $\mathcal{M}_2$ is a monotone class on E.
Since $\mathcal{A} \subset \mathcal{M}_2$ as shown above, by the definition of $\mathcal{M}$ as the smallest monotone class containing $\mathcal{A}$, we have $\mathcal{M} \subset \mathcal{M}_2$, so that $B \setminus A \in \mathcal{M}$ for any $B, A \in \mathcal{M}$.

It now follows that $\mathcal{M}$ is a λ-system on E containing $\mathcal{A}$. Since $\mathcal{A}$ is closed under finite intersections and thus a π-system, by the $\pi - \lambda$ theorem the σ-algebra generated by $\mathcal{A}$ is contained in $\mathcal{M}$.
As such, the smallest montone class $\mathcal{M}$ containing $\mathcal{A}$ is exactly the σ-algebra generated by $\mathcal{A}$, which is the smallest σ-algebra containing $\mathcal{A}$.
Q.E.D.

2.2 Measurable Functions

2.2.1 Measurability and Characterizations

Let $(E,\mathcal{E})$, $(F,\mathcal{F})$ be measurable spaces, and $f: E \to F$ a function that maps E into F. f is said to be a measurable function relative to the σ-algebras $\mathcal{E}$ and $\mathcal{F}$ if:

$$\text{For any } A \in \mathcal{F}, \text{ the inverse image } f^{-1}(A) \in \mathcal{E}.$$

We denote this relation succinctly by $f \in \mathcal{E}/\mathcal{F}$.
Because we will encounter functions $f: E \to [-\infty,+\infty]$ very often, we write $f \in \mathcal{E}$ for $f \in \mathcal{E}/\mathcal{B}([-\infty,+\infty])$. If f is nonnegative valued, then we write $f \in \mathcal{E}_+$.

The following result simplifies the criteria for measurability by relying on generating sets.

Lemma 2.5 Let $\mathcal{F}_0$ be a collection of subsets of F that generates $\mathcal{F}$. Then, a function $f: E \to F$ is measurable relative to $\mathcal{E}$ and $\mathcal{F}$ if and only if $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{F}_0$.

Proof) The necessity part follows immediately, since $\mathcal{F}_0 \subset \mathcal{F}$ and $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{F}$ if $f \in \mathcal{E}/\mathcal{F}$.

To show sufficiency, suppose that $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{F}_0$, and define

$$\mathcal{M} = \{A \subset F \mid f^{-1}(A) \in \mathcal{E}\};$$

$\mathcal{M}$ contains $\mathcal{F}_0$. We now show that $\mathcal{M}$ is a σ-algebra on F:

i) $f^{-1}(F) = E \in \mathcal{E}$, so $F \in \mathcal{M}$.

ii) For any $A \in \mathcal{M}$,

$$f^{-1}(F \setminus A) = f^{-1}(F) \setminus f^{-1}(A) = E \setminus f^{-1}(A) \in \mathcal{E}$$

because $f^{-1}(A) \in \mathcal{E}$ and $\mathcal{E}$ is closed under complements. This implies that $A^c = F \setminus A \in \mathcal{M}$.

iii) For any countable collection $\{A_n\}_{n \in N_+} \subset \mathcal{M}$, letting $A = \bigcup_n A_n$,

$$f^{-1}(A) = \bigcup_n f^{-1}(A_n) \in \mathcal{E}$$

becauase each $f^{-1}(A_n) \in \mathcal{E}$ and $\mathcal{E}$ is closed under countable unions.

It follows that $\mathcal{M}$ is a σ-algebra on F containing $\mathcal{F}_0$. Since $\mathcal{F} = \sigma\mathcal{F}_0$ is the smallest σ-algebra containing $\mathcal{F}_0$, we have $\mathcal{F} \subset \mathcal{M}$. In other words, $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{F}$, so by definition, $f \in \mathcal{E}/\mathcal{F}$.
Q.E.D.

The powerful result above gives us the following characterization of measurability for functions whose target space is a Borel space, and shows that continuous functions are always measurable:

Corollary to Lemma 2.5 The following hold true:

i) Let $(E, \mathcal{E})$ be a measurable space and $(F, \mathcal{B}(F, \tau))$ a Borel space. Then, a function $f : E \to F$ is measurable relative to $\mathcal{E}$ and $\mathcal{B}(F, \tau)$ if and only if $f^{-1}(A) \in \mathcal{E}$ for any $A \in \tau$.

ii) Let (E, τ) and (F, s) be topological spaces. Then, any function $f : E \to F$ continuous relative to τ and s is also measurable relative to $\mathcal{B}(E, \tau)$ and $\mathcal{B}(F, s)$.

Proof) i) This follows immediately from lemma 2.5 because τ generates the σ-algebra $\mathcal{B}(F, \tau)$.

ii) Let (E, τ) and (F, s) be topological spaces and $f : E \to F$ a function continuous relative to τ and s. By definition, $f^{-1}(A) \in \tau$ for any $A \in s$. Since $\sigma\tau = \mathcal{B}(E, \tau)$, this means that $f^{-1}(A) \in \mathcal{B}(E, \tau)$ for any $A \in s$, and by result i), this indicates that $f \in \mathcal{E}/\mathcal{F}$.

Q.E.D.

Furthermore, lemma 2.5 implies that, for any measurable space $(E, \mathcal{E})$ and a real function $f : E \to \mathbb{R}$, $f \in \mathcal{E}/\mathcal{B}(\mathbb{R})$ if and only if $f^{-1}((a, b)) \in \mathcal{E}$ for any open interval (a, b) with rational endpoints, since the base of such open intervals was shown to generate the standard topology on $\mathbb{R}$.
Likewise, for any function $f : E \to [-\infty, +\infty]$ that takes values on the extended real line, $f \in \mathcal{E}$ if and only if

$$f^{-1}([-\infty, a)), f^{-1}((a, +\infty]), f^{-1}((a, b)) \in \mathcal{E}$$

for any rational $a, b \in \mathbb{Q}$. This characterization of measurability can be further simplified, as shown below:

Theorem 2.6 Let $(E, \mathcal{E})$ be a measurable space, and let $f : E \to [-\infty, +\infty]$ be a function. Then, the following statements are equivalent:

i) f is $\mathcal{E}$-measurable

ii) $f^{-1}([-\infty, a)) \in \mathcal{E}$ for any $a \in \mathbb{Q}$

iii) $f^{-1}([a, +\infty]) \in \mathcal{E}$ for any $a \in \mathbb{Q}$

iv) $f^{-1}([-\infty, a]) \in \mathcal{E}$ for any $a \in \mathbb{Q}$

v) $f^{-1}((a, +\infty]) \in \mathcal{E}$ for any $a \in \mathbb{Q}$

Proof) We show that each statement implies the next.

i) $\to$ *ii)* Suppose $f \in \mathcal{E}$. Then, because $[-\infty, a) \in \tau_{[-\infty,+\infty]} \subset \mathcal{B}([-\infty, +\infty])$ for any $a \in \mathbb{Q}$, by the definition of measurability $f^{-1}([-\infty, a)) \in \mathcal{E}$ for any $a \in \mathbb{Q}$.

ii) $\leftrightarrow$ *iii)* Suppose $f^{-1}([-\infty, a)) \in \mathcal{E}$ for any $a \in \mathbb{Q}$. Then, for any $a \in \mathbb{Q}$,

$$\begin{aligned} f^{-1}([a, +\infty]) &= f^{-1}\left([-\infty, +\infty] \setminus [-\infty, a)\right) \\ &= f^{-1}\left([-\infty, +\infty]\right) \setminus f^{-1}([-\infty, a)) = E \setminus f^{-1}([-\infty, a)) = f^{-1}([-\infty, a))^c \in \mathcal{E}, \end{aligned}$$

where the last inclusion follows because $\mathcal{E}$ is closed under complements.
Conversely, suppose that *iii)* holds. Because $[-\infty, a) = [-\infty, +\infty] \setminus [a, +\infty]$ for any $a \in \mathbb{Q}$, by the same reasoning as above,

$$f^{-1}([-\infty, a)) \in \mathcal{E}$$

and thus *ii)* holds.

ii) $\to$ *iv)* Suppose *ii)* holds. Then, for any $a \in \mathbb{Q}$, because

$$[-\infty, a] = \bigcap_{n \in N_+} \left[-\infty, a + \frac{1}{n}\right),$$

we can see that

$$f^{-1}([-\infty, a]) = f^{-1}\left(\bigcap_{n \in N_+} \left[-\infty, a + \frac{1}{n}\right)\right) = \bigcap_{n \in N_+} f^{-1}\left(\left[-\infty, a + \frac{1}{n}\right)\right) \in \mathcal{E},$$

where the last inclusion holds because each $f^{-1}\left(\left[-\infty, a + \frac{1}{n}\right)\right)$ is measurable by hypothesis and $\mathcal{E}$ is closed under countable intersections.
This holds for any $a \in \mathbb{Q}$, so *iv)* holds.

iv) $\leftrightarrow$ *v)* Note that the sets of the form $[-\infty,a]$ and $(a,+\infty]$ are complements of one another. Therefore, the closedness of σ-algebras under complements indicate that, if *iv)* or *v)* hold, then the other holds as well.

v) $\rightarrow$ *i)* Suppose that $f^{-1}((a,+\infty])\in\mathcal{E}$ for any $a\in\mathbb{Q}$. Then, we showed above that $f^{-1}([-\infty,a])\in\mathcal{E}$ for any $a\in\mathbb{Q}$ as well.
Choose any $a\in\mathbb{Q}$, and note that

$$[-\infty,a)=\bigcup_{n\in N_+}\left[-\infty,a-\frac{1}{n}\right];$$

as such,

$$f^{-1}([-\infty,a))=f^{-1}\left(\bigcup_{n\in N_+}\left[-\infty,a-\frac{1}{n}\right]\right)=\bigcup_{n\in N_+}f^{-1}\left(\left[-\infty,a-\frac{1}{n}\right]\right)\in\mathcal{E},$$

where the last inclusion holds because each $f^{-1}\left(\left[-\infty,a-\frac{1}{n}\right]\right)$ is measurable and $\mathcal{E}$ is closed under countable unions.
Finally, for any $a,b\in\mathbb{Q}$ such that $a<b$, because

$$(a,b)=[-\infty,b)\setminus[a,+\infty],$$

we have

$$f^{-1}((a,b))=f^{-1}([-\infty,b)\setminus[a,+\infty])=f^{-1}([-\infty,b))\setminus f^{-1}([a,+\infty])\in\mathcal{E}.$$

We have thus shown that $f^{-1}(A)\in\mathcal{E}$ for any $A\in\tau_{[-\infty,+\infty]}$ contained in the countable base

$$\bar{\mathcal{B}}=\{[-\infty,a)\mid a\in\mathbb{Q}\}\bigcup\{(a,+\infty]\mid a\in\mathbb{Q}\}\bigcup\{(a,b)\mid a,b\in\mathbb{Q}\}.$$

Since $\bar{\mathbb{B}}$ is a countable base generating the standard topology $\tau_{[-\infty,+\infty]}$ on $[-\infty,+\infty]$ by theorem 1.5, lemma 2.2 shows us that $\bar{\mathbb{B}}$ generates the Borel σ-algebra $\mathcal{B}([-\infty,+\infty])$, and therefore, by lemma 2.5, f is measurable relative to $\mathcal{E}$ and $\mathcal{B}([-\infty,+\infty])$.

Q.E.D.

The workhorse theorem proved above allows us to easily prove many statements concerning the measurability of numerical functions, or functions taking values in $[-\infty,+\infty]$.

The following is a corollary to the above theorem that extends the above criteria to real-valued functions.

Corollary to Theorem 2.6 Let $(E,\mathcal{E})$ be a measurable space, and let $f:E\to\mathbb{R}$ a function. Then, the following statements are equivalent:

i) f is $\mathcal{E}$-measurable

ii) $f^{-1}((-\infty,a))\in\mathcal{E}$ for any $a\in\mathbb{Q}$

iii) $f^{-1}([a,+\infty))\in\mathcal{E}$ for any $a\in\mathbb{Q}$

iv) $f^{-1}((-\infty,a])\in\mathcal{E}$ for any $a\in\mathbb{Q}$

v) $f^{-1}((a,+\infty))\in\mathcal{E}$ for any $a\in\mathbb{Q}$

Proof) Let $\bar{f}:E\to[-\infty,+\infty]$ be the extension of f to the extended real line. If *i)* holds, then *ii)* to *v)* hold because $(-\infty,a),[a,+\infty),(-\infty,a],(a,+\infty)\in\tau_{\mathbb{R}}\subset\mathcal{B}(\mathbb{R})$ for any $a\in\mathbb{Q}$. Suppose *ii)* holds. Then, for any $a\in\mathbb{Q}$,

$$\bar{f}^{-1}([-\infty,a))=f^{-1}((-\infty,a))\in\mathcal{E},$$

and by theorem 2.6, $\bar{f}$ is $\mathcal{E}$-measurable. By definition,

$$\bar{f}^{-1}(A)\in\mathcal{E}$$

for any $A\in\mathcal{B}([-\infty,+\infty])$. Since $(a,b)\in\mathcal{B}([-\infty,+\infty])$ for any $a,b\in\mathbb{Q}$ because $\mathcal{B}([-\infty,+\infty])$ is generated by the order topology on $[-\infty,+\infty]$, we have

$$f^{-1}((a,b))=\bar{f}^{-1}((a,b))\in\mathcal{E}$$

for any $a,b\in\mathbb{Q}$. It follows from lemma 2.5 that $f\in\mathcal{E}/\mathcal{B}(\mathbb{R})$ and thus that *i)* holds. Similar arguments show that *iii)* to *v)* also imply that *i)* holds.
Q.E.D.

The above reveals that, for any $f:E\to\mathbb{R}$, if $f\in\mathcal{E}/\mathcal{B}(\mathbb{R})$ then the extension $\bar{f}:E\to[-\infty,+\infty]$ of f to the extended real number system must be $\mathcal{E}$-measurable. Conversely, if $f:E\to[-\infty,+\infty]$ is real valued and $\mathcal{E}$-measurable, then its restriction $\tilde{f}:E\to\mathbb{R}$ to $\mathbb{R}$ is in $\mathcal{E}/\mathcal{B}(\mathbb{R})$. Thus, we may identify measurability of real-valued functions regardless of whether its target space is the real line or the extended real-number system; for this reason, we will also denote $f\in\mathcal{E}$, or call f $\mathcal{E}$-measurable, if f is a function with target space $\mathbb{R}$ that is measurable with respect to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R})$.

2.2.2 Preservation of Measurability under Various Operations

Here we show that measurability is preserved across compositions and simple arithmetic operations.

Theorem 2.7 The following hold true:

i) Let $(E,\mathcal{E})$, $(F,\mathcal{F})$, and $(G,\mathcal{G})$ be measurable spaces, and let $f : E \to F$, $g : F \to G$ be functions.
If $f \in \mathcal{E}/\mathcal{F}$ and $g \in \mathcal{F}/\mathcal{G}$, then the function $h : E \to G$ defined as $h = g \circ f$ is measurable relative to $\mathcal{E}$ and $\mathcal{G}$.

ii) Let $(E,\mathcal{E})$ be a measurable space, and $f_1,\cdots,f_n : E \to \mathbb{R}$ measurable functions relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R})$. Let $\Phi : \mathbb{R}^n \to \mathbb{R}$ be measurable relative to $\mathcal{B}(\mathbb{R}^n)$ and $\mathcal{B}(\mathbb{R})$.
Then, the function $h : E \to \mathbb{R}$ defined as

$$h(x) = \Phi(f_1(x),\cdots,f_n(x))$$

for any $x \in E$ is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R}^n)$.

iii) Let $(E,\mathcal{E})$ be a measurable space, and suppose $f,g : E \to \mathbb{R}$ are real valued functions on E. If $f,g \in \mathcal{E}/\mathcal{B}(\mathbb{R})$, then $f+g$ and fg are measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R})$ as well.

iv) Let $(E,\mathcal{E})$ be a measurable space, and suppose $f : E \to [-\infty,+\infty]$ is $\mathcal{E}$-measurable. Then, for any $c \in \mathbb{R}$, cf is also $\mathcal{E}$-measurable.

v) Let $(E,\mathcal{E})$ be a measurable space, and suppose $f,g : E \to [-\infty,+\infty]$ are $\mathcal{E}$-measurable. Then, the sets

$$\{f \geq g\}, \quad \{f \leq g\}, \quad \{f > g\}, \quad \{f < g\}, \quad \{f = g\}, \quad \{f \neq g\}$$

are all $\mathcal{E}$-measurable sets.

Proof) i) Suppose that $f \in \mathcal{E}/\mathcal{F}$ and $g \in \mathcal{F}/\mathcal{G}$. Then, for any $A \in \mathcal{G}$,

$$h^{-1}(A) = f^{-1}(g^{-1}(A)) \in \tau$$

because $g^{-1}(A) \in \mathcal{F}$ and f is measurable. By definition, $h \in \mathcal{E}/\mathcal{G}$.

ii) We first show that the function $f : E \to \mathbb{R}^n$ defined as

$$f(x) = (f_1(x),\cdots,f_n(x))$$

for any $x \in E$ is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R}^n)$. Letting τ_n^e be the euclidean topology on $\mathbb{R}^n$, we showed in theorem 1.9 that

$$\tau_n^e = \underbrace{\tau_{\mathbb{R}} \times \cdots \times \tau_{\mathbb{R}}}_{n} = \tau_{\mathbb{R}}^n.$$

In addition, $\mathbb{B} = \{(a,b) \in a, b \in \mathbb{Q}\}$ is a countable base that generates $\tau_{\mathbb{R}}$ (by theorem 1.5), and as such, by lemma 1.8, the collection

$$\mathbb{B}^n = \{B_1 \times \cdots \times B_n \mid B_i \in \mathbb{B} \text{ for any } 1 \leq i \leq n\}$$

of open rectangles on $\mathbb{R}^n$ is a countable base generating $\tau_{\mathbb{R}}^n = \tau_n^e$.
Therefore, by lemma 2.2, $\mathbb{B}^n$ generates the σ-algebra $\mathcal{B}(\mathbb{R}^n)$. For any $B = B_1 \times \cdots \times B_n \in \mathbb{B}^n$,

$$f^{-1}(B) = \bigcap_{i=1}^{n} f_i^{-1}(B_i) \in \mathcal{E},$$

where the last inclusion holds because each $f_i^{-1}(B_i) \in \mathcal{E}$ by the measurability of f_i and $\mathcal{E}$ is closed under finite intersections.
This holds for any $B \in \mathbb{B}^n$, so by lemma 2.5, $f \in \mathcal{E}/\mathcal{B}(\mathbb{R}^n)$.

The remaining step is to show that the function $h : E \to \mathbb{R}$ defined as

$$h(x) = \Phi(f_1(x), \cdots, f_n(x))$$

for any $x \in E$ is $\mathcal{E}$-measurable. However, because we can write $h = \Phi \circ f$, where $\Phi \in \mathcal{B}(\mathbb{R}^n)/\mathcal{B}(\mathbb{R})$ by hypothesis and we showed that $f \in \mathcal{E}/\mathcal{B}(\mathbb{R}^n)$, it follows from result *i*) that $h \in \mathcal{E}/\mathcal{B}(\mathbb{R})$.

iii) Note that the functions $\Phi_1, \Phi_2 : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$\Phi_1(x) = x_1 + x_2 \quad \text{and} \quad \Phi_2(x) = x_1 x_2$$

for any $x = (x_1, x_2) \in \mathbb{R}^2$ are continuous relative to the euclidean topologies on $\mathbb{R}^2$ and $\mathbb{R}$; this is easily shown using elementary analytical machinery.

As such, $\Phi_1, \Phi_2 \in \mathcal{B}(\mathbb{R}^2)/\mathcal{B}(\mathbb{R})$ by the corollary to lemma 2.5, and from result *ii*), we can see that

$$f + g = \Phi_1(f, g) \quad \text{and} \quad fg = \Phi_2(f, g)$$

are measurable relative to $\mathcal{E}$ and $\mathbb{R}$.

iv) If $c = 0$, then $cf(x) = 0$ for any $x \in E$, so $cf \in \mathcal{E}$.
Now let $c > 0$. Then, for any $r \in \mathbb{Q}$, by the denseness of $\mathbb{Q}$ in $\mathbb{R}$, we have

$$\{cf < r\} = \{f < \frac{r}{c}\} = \bigcup_{q < r/c,\ q \in \mathbb{Q}} f^{-1}([-\infty, q));$$

this union is a countable union of measurable sets, so $\{cf < r\} \in \mathcal{E}$ and $cf \in \mathcal{E}$.
If $c < 0$, then $\{cf > r\} = \{f < \frac{r}{c}\}$, so by the same process as above, we can show that $cf \in \mathcal{E}$.

v) By the denseness of $\mathbb{Q}$ in $\mathbb{R}$, note that

$$\begin{aligned}\{f > g\} &= \bigcup_{q > r,\ q, r \in \mathbb{Q}} \{x \in E \mid f(x) > q > r > g(x)\} \\ &= \bigcup_{q > r,\ q, r \in \mathbb{Q}} [\{x \in E \mid f(x) > q\} \cap \{x \in E \mid r > g(x)\}].\end{aligned}$$

Beacause $\{x \in E \mid f(x) > q\}, \{x \in E \mid r > g(x)\} \in \mathcal{E}$ for any $q, r \in \mathbb{Q}$ by the measurability of f, g, and the union and intersection above are over a countable collection of sets, it follows that $\{f > g\} \in \mathcal{E}$.
Likewise, $\{f < g\} = \{-f > -g\} \in \mathcal{E}$ because $-f, -g \in \mathcal{E}$ by *iv*).
In addition, $\{f \leq g\} = \{f > g\}^c \in \mathcal{E}$ and $\{f \geq g\} = \{f < g\}^c \in \mathcal{E}$.
Finally, $\{f = g\} = \{f \leq g\} \setminus \{f < g\} \in \mathcal{E}$, and $\{f \neq g\} = \{f = g\}^c \in \mathcal{E}$ as well.

Q.E.D.

2.2.3 Preservation of Measurability under Limits

One of the main advantages of working with measurable functions instead of continuous functions is that the property of measurability, unlike continuity, is preserved under pointwise limits as well as uniform limits, whereas the pointwise limit of continuous functions may be discontinuous. This is formally shown below:

Theorem 2.8 Let $(E, \mathcal{E})$ be a measurable space. The following hold true:

i) Let $\{f_n\}_{n\in N_+}$ be a sequence of $\mathcal{E}$-measurable functions. Then, $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are $\mathcal{E}$-measurable numerical functions.

ii) Let $\{f_n\}_{n\in N_+}$ be a sequence of $\mathcal{E}$-measurable functions with a pointwise limit $f = \lim_{n\to\infty} f_n$ taking values in $[-\infty, +\infty]$. Then, f is $\mathcal{E}$-measurable.

Proof) i) By definition,

$$\limsup_{n\to\infty} f_n = \inf_{n\in N_+}\left(\sup_{k\geq n} f_k\right)$$

and

$$\liminf_{n\to\infty} f_n = \sup_{n\in N_+}\left(\inf_{k\geq n} f_k\right).$$

Therefore, to establish the measurability of the lim sup and lim inf functions, we must first establish the measurability of supremums and infimums of functions.

For any $n \in N_+$, define $g_n : E \to [-\infty, +\infty]$ as

$$g_n(x) = \sup_{k\geq n} f_k(x)$$

for any $x \in E$. For any $a \in \mathbb{Q}$,

$$g_n^{-1}([-\infty, a]) = \{x \in E \mid \sup_{k\geq n} f_k(x) \leq a\} = \bigcap_{k\geq n}\{x \in E \mid f_k(x) \leq a\} = \bigcap_{k\geq n} f_k^{-1}([-\infty, a]).$$

Because each $f_k \in \mathcal{E}$, the sets $f_k^{-1}([-\infty, a]) \in \mathcal{E}$ by theorem 2.6, and since $\mathcal{E}$ is closed under countable intersections, $g_n^{-1}([-\infty, a]) \in \mathcal{E}$ as well. This holds for any $a \in \mathbb{Q}$, so by theorem 2.4, $g_n \in \mathcal{E}$.

This holds for any $n \in N_+$, so $\{g_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable numerical functions. Then, $f = \limsup_{n\to\infty} f_n = \inf_{n\in N_+} g_n$; for any $a \in \mathbb{Q}$,

$$f^{-1}([a, +\infty]) = \{x \in E \mid \inf_{n\in N_+} g_n \geq a\} = \bigcap_{n\in N_+}\{x \in E \mid g_n(x) \geq a\} = \bigcap_{n\in N_+} g_n^{-1}([a, +\infty]).$$

Each $g_n^{-1}([a,+\infty]) \in \mathcal{E}$ and $\mathcal{E}$ is closed under countable intersections, so $f^{-1}([a,+\infty]) \in \mathcal{E}$. This holds for any $a \in \mathbb{Q}$, so by theorem 2.6, $f \in \mathcal{E}$, or in other words, $\limsup_{n\to\infty} f_n$ is measurable.

As for $\liminf_{n\to\infty} f_n$, since

$$\liminf_{n\to\infty} f_n = -\inf_{n\in N_+}\left(\sup_{k\geq n}(-f_k)\right) = -\limsup_{n\to\infty}(-f_n),$$

and $\{-f_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable functions, the result shown above implies that $\liminf_{n\to\infty} f_n$ is measurable as well.

ii) If $\{f_n\}_{n\in N_+}$ be a sequence of $\mathcal{E}$-measurable functions witha pointwise limit $f : E \to [-\infty,+\infty]$, for any $x \in E$ we have

$$f(x) = \lim_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x) = \limsup_{n\to\infty} f_n(x),$$

where the last two equalities follow from the fact that the sequence $\{f_n(x)\}_{n\in N_+}$ converges. Therefore,

$$f = \liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n;$$

the latter two functions were shown to be measurable, so it follows that f is $\mathcal{E}$-measurable as well.

Q.E.D.

2.2.4 The Positive and Negative Parts of a Function

Let $f : E \to [-\infty, +\infty]$, and define $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$. Then, for any $x \in E$, if $f(x) \geq 0$, then $f(x) = f^+(x)$ and $f^-(x) = 0$, so that $f(x) = f^+(x) - f^-(x)$, while if $f(x) < 0$, then $f(x) = -f^-(x)$ and $f^+(x) = 0$, so that $f(x) = f^+(x) - f^-(x)$ once again. Therefore, we can decompose f into its positive and negative parts on the entire set E:

$$f = f^+ - f^-.$$

Note that f^+ and f^- both non-negative valued, that is, they take values on $[0, +\infty]$.

The positive and negative parts of a function have useful measurability and minimality properties: these are stated below.

Lemma 2.9 Let $(E, \mathcal{E})$ be a measurable space, and $f : E \to [-\infty, +\infty]$ a numerical function. The following hold true:

i) f is $\mathcal{E}$-measurable if and only if the positive and negative parts f^+ and f^- of f are $\mathcal{E}$-measurable.

ii) Let $f = g - h$ for some $g, h : E \to [-\infty, +\infty]$ that are non-negative valued. Then, $f^+ \leq g$ and $f^- \leq h$.

Proof) i) Suppose f is $\mathcal{E}$-measurable. The function $g : E \to [-\infty, +\infty]$ defined as $g(x) = 0$ for any $x \in E$ is $\mathcal{E}$-measurable, since

$$g^{-1}([-\infty, a)) = \begin{cases} E & \text{if } a > 0 \\ \emptyset & \text{if } a \leq 0 \end{cases} \in \mathcal{E}$$

for any $a \in \mathbb{Q}$.
Now we can write $f^+ = \max(f, g)$ and $f^- = \min(f, g)$; it follows from the preservation of measurability under supremums and infimums and the measurability of f and g that $f^+, f^- \in \mathcal{E}$.
Conversely, suppose that f^+, f^- are $\mathcal{E}$-measurable. Choose any $a \in \mathbb{Q}$, and suppose $a > 0$. If $f(x) < a$ for some $x \in E$, then because $0 < a$ as well, $f^+(x) = \max(f(x), 0) < a$. Conversely, if $f^+(x) < a$ for some $x \in E$, then because $f(x) \leq f^+(x)$, $f(x) < a$ as well. This equivalency can be written as

$$\begin{aligned} f^{-1}([-\infty, a)) &= \{x \in E \mid f(x) < a\} \\ &= \{x \in E \mid f^+(x) < a\} = (f^+)^{-1}([-\infty, a)) \in \mathcal{E}, \end{aligned}$$

where the last inclusion holds because f^+ is $\mathcal{E}$-measurable.
Now suppose $a \leq 0$. If $f(x) < a$, then because $f(x) < a \leq 0$, $\min(f(x), 0) = f(x) < a$ and $f^-(x) = -\min(f(x), 0) = -f(x) > -a$. Conversely, if $f^-(x) > -a$, then because

$\min(f(x),0) < a \leq 0$, $\min(f(x),0) = f(x)$ and $f(x) < a$. This equivalency can be written as

$$\begin{aligned}f^{-1}([-\infty,a)) &= \{x \in E \mid f(x) < a\} \\ &= \{x \in E \mid f^-(x) > -a\} = (f^-)^{-1}((-a,+\infty]) \in \mathcal{E},\end{aligned}$$

where the last inclusion holds because f^- is $\mathcal{E}$-measurable.
Therefore, $f^{-1}([-\infty,a)) \in \mathcal{E}$ for any $a \in \mathbb{Q}$, and by the characterization of measurability for numerical functions, f is $\mathcal{E}$-measurable.

ii) Suppose $f = g - h$ for some $g,h : E \to [-\infty,+\infty]$ that are non-negative valued. Then,

$$f^+ = \max(f,0) \leq f = g - h \leq g$$

on E, where the last inequality follows from the non-negativity of h. Likewise,

$$f^- = -\min(f,0) \leq -f = h - g \leq h$$

on E by the non-negativity of g.

Q.E.D.

Due to the ubiquity of non-negative measurable functions, from now on we denote a non-negative measurable function f by $f \in \mathcal{E}_+$.

2.2.5 Simple Functions

Let $(E,\mathcal{E})$ be a measurable space. A simple function $f:E\to\mathbb{R}$ is defined as a function that maps into a finite subset of $[0,+\infty)$. Letting $f:E\to\mathbb{R}$ be a real function mapping into the finite set $\{\alpha_1,\cdots,\alpha_n\}\subset[0,+\infty)$ of distinct elements, for any $1\le i\le n$ we can define the inverse image $A_i=f^{-1}(\{\alpha_i\})$; then, f can be written as

$$f=\sum_{i=1}^{n}\alpha_i\cdot I_{A_i}.$$

Note that $A_1,\cdots,A_n\subset E$ are disjoint, since $x\in A_i\cap A_j$ for some $i\neq j$ indicates that $f(x)=\alpha_i=\alpha_j$, which contradicts the fact that $\alpha_i\neq\alpha_j$. A representation of a simple function in the form above, or the linear combination of indicators of disjoint sets whose union is E with distinct coefficients, is referred to as the canonical form of f.
Conversely, suppose that

$$f=\sum_{i=1}^{n}\alpha_i\cdot I_{A_i}$$

for some $\alpha_1,\cdots,\alpha_n\in[0,+\infty)$ and $A_1,\cdots,A_n\subset E$, where $\alpha_1,\cdots,\alpha_n$ may not be distinct and $A_1,\cdots,A_n$ not disjoint. Nevertheless, f takes values in a finite subset of $\{\alpha_1,\cdots,\alpha_n\}$, so it is a simple function.

Let f be a simple function on E with canonical form

$$f=\sum_{i=1}^{n}\alpha_i\cdot I_{A_i}.$$

It can be easily shown that f is $\mathcal{E}$-measurable if and only if $A_1,\cdots,A_n\in\mathcal{E}$:

- If $f\in\mathcal{E}$, then for any $1\le i\le n$, because $\{\alpha_i\}\in\mathcal{B}(\mathbb{R})$, we have $A_i=f^{-1}(\{\alpha_i\})\in\mathcal{E}$ by the definition of measurability.

- If $A_1,\cdots,A_n\in\mathcal{E}$, then for any $A\in\mathcal{B}(\mathbb{R})$, because $f^{-1}\left(\mathbb{R}\setminus\{\alpha_1,\cdots,\alpha_n\}\right)=\emptyset$, we have

  $$f^{-1}(A)=f^{-1}\left(A\cap\{\alpha_1,\cdots,\alpha_n\}\right)=\bigcup_{i=1}^{n}f^{-1}\left(\{\alpha_i\}\cap A\right).$$

  Defining $B_i=\{\alpha_i\}\cap A$ for $1\le i\le n$, if $B_i=\emptyset$, then $f_{-1}(B_i)=\emptyset\in\mathcal{E}$, while if $B_i\neq\emptyset$, then $B_i=\{\alpha_i\}$ and $f^{-1}(B_i)=f^{-1}(\{\alpha_i\})=A_i\in\mathcal{E}$. Therefore, $f^{-1}(B_1),\cdots,f^{-1}(B_n)\in\mathcal{E}$ and $f^{-1}(A)\in\mathcal{E}$.
  This holds for any $A\in\mathcal{B}(\mathbb{R})$, so f is $\mathcal{E}$-measurable.

We now prove a very powerful result that every measurable numerical function is the pointwise limit of an increasing sequence of simple functions.

Theorem 2.10 (Approximation by Simple Functions)
Let $(E, \mathcal{E})$ be a measurable space, and $f : E \to [-\infty, +\infty]$ a non-negative $\mathcal{E}$-measurable function. Then, there exists an increasing sequence $\{s_n\}_{n \in N_+}$ of $\mathcal{E}$-measurable simple functions such that $s_n \nearrow f$ pointwise as $n \to \infty$.

Proof) For any $n \in N_+$, define $s_n : E \to [0, +\infty)$ as

$$s_n(x) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot I_{f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right)}(x) + n \cdot I_{f^{-1}([n,+\infty])}(x)$$

for any $x \in E$. s_n takes values in the finite set $\{\frac{k-1}{2^n} \mid 1 \leq k \leq 2^n + 1\}$, so by definition, s_n is a simple function. Furthermore, for any $1 \leq k \leq n2^n$, the set $\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right) \subset [0, +\infty]$ is a Borel set, so by measurability, $f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right) \in \mathcal{E}$. Thus, as we showed above, $s_n \in \mathcal{E}_+$.

We must now show that $s_n \leq s_{n+1}$ on E for any $n \in N_+$. Choose any $n \in N_+$ and $x \in E$, and denote $y = f(x) \in [0, +\infty]$. We now study three distinct cases:

- If $y \geq n+1$, then $s_n(x) = n < n+1 = s_{n+1}(x)$.
- If $n \leq y < n+1$, then letting $y \in \left[\frac{k-1}{2^{n+1}}, \frac{k}{2^{n+1}}\right)$ for some $n \cdot 2^{n+1} + 1 \leq k \leq (n+1)2^{n+1}$,

$$s_{n+1}(x) = \frac{k-1}{2^{n+1}} \geq \frac{n \cdot 2^{n+1} + 1 - 1}{2^{n+1}} = n = s_n(x).$$

- Finally, if $y < n$, then letting $y \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$ for some $1 \leq k \leq n2^n$, we can see that

$$y \in \left[\frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}}\right) \bigcup \left[\frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}}\right),$$

  so that

$$s_{n+1}(x) = \frac{2k-2}{2^{n+1}} \text{ or } \frac{2k-1}{2^{n+1}}.$$

  In any case, $s_{n+1}(x)$ is larger than or equal to $s_n(x) = \frac{k-1}{2^n}$.

Therefore, in any case, $s_{n+1}(x) \geq s_n(x)$, so $s_n \leq s_{n+1}$ on E.

Finally, we must show that $s_n \to f$ pointwise as $n \to \infty$. Choose any $x \in E$, $\varepsilon > 0$, and denote $y = f(x) \in [0, +\infty]$.
If $y = +\infty$, then $s_n(x) = n$ for any $n \in N_+$ and $s_n(x) \to +\infty = y = f(x)$ as $n \to \infty$.
On the other hand, suppose $y < +\infty$. Then, there exists a natural number $M \in N_+$ such that $y < M$, and for any $n \geq M$, there exists a $1 \leq k \leq n2^n$ such that $y \in \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$.
As such,

$$s_n(x) = \frac{k-1}{2^n} \leq y = f(x) < \frac{k}{2^n} = \frac{1}{2^n} + s_n(x),$$

so that

$$|s_n(x) - f(x)| = f(x) - s_n(x) < \frac{1}{2^n}.$$

Because $\frac{1}{2^k} \to 0$ as $k \to \infty$, there exists an $N \geq M$ such that $\frac{1}{2^n} < \varepsilon$ for any $n \geq N$; thus, for any $n \geq N$, $n \geq M$ implies that

$$|s_n(x) - f(x)| < \frac{1}{2^n} < \varepsilon.$$

Such an $N \in N_+$ holds for any $\varepsilon > 0$, so by definition

$$\lim_{n\to\infty} s_n(x) = f(x).$$

In any case, $s_n(x) \to f(x)$; this holds for any $x \in E$, so by definition $s_n \to f$ pointwise as $n \to \infty$.
Q.E.D.

Simple functions are also useful in that the sum or product of simple functions is also a simple function:

Lemma 2.11 Let $(E, \mathcal{E})$ be a measurable space, and $f, g \in \mathcal{E}_+$ measurable simple functions. Then, $f + g$ and fg are also measurable simple functions.

Proof) Let the canonical forms of f and g be given as

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i} \quad \text{and} \quad g = \sum_{i=1}^{m} \beta_i \cdot I_{B_i},$$

where $\bigcup_i A_i = \bigcup_i B_i = E$. Define $C_{ij} = A_i \cap B_j$ for any $1 \leq i \leq n$ and $1 \leq j \leq m$. Due to the disjoint nature of $B_1, \cdots, B_m$,

$$I_{A_i} = I_{A_i \cap \left(\bigcup_{j=1}^{m} B_j\right)} = \sum_{j=1}^{m} I_{A_i \cap B_j},$$

and

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i \cap \left(\bigcup_j B_j\right)} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \cdot I_{C_{ij}}.$$

Likewise, for g,

$$g = \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j \cdot I_{C_{ij}}.$$

Therefore,

$$f+g=\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i+\beta_j)\cdot I_{C_{ij}}.$$

Each $\alpha_i+\beta_j \in [0,+\infty)$, and $C_{ij}\subset E$, so $f+g$ is a simple function. In addition, f and g are real valued measurable functions, and theorem 2.7 tells us that the sum of measurable functions is measurable, so $f+g$ is a measurable simple function.

The case for fg is easier. Note that

$$fg=\left(\sum_{i=1}^{n}\alpha_i\cdot I_{A_i}\right)\left(\sum_{i=1}^{m}\beta_i\cdot I_{B_i}\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i\beta_j)\cdot I_{A_i\cap B_j}=\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i\beta_j)\cdot I_{C_{ij}}.$$

It follows immediately that fg is a simple function, and because the measurability of real valued functions is preserved under products, fg is also a measurable simple function.
Q.E.D.

A useful corollary of the above results is that the sum of non-negative functions is also measurable; note that the functions in questions do not need to be real valued, and can take $+\infty$ as a value.

Corollary to Theorem 2.10 and Lemma 2.11 Let $(E,\mathcal{E})$ be a measurable space, and $f,g\in\mathcal{E}_+$ non-negative measurable functions. Then, $f+g$ and fg are also non-negative measurable functions.

Proof) By theorem 2.10, there exist sequences $\{s_n\}_{n\in N_+}$ and $\{h_n\}_{n\in N_+}$ of simple measurable functions increasing to f and g respectively. The sum of simple functions is also a simple function, and measurability is preserved over the sum of real-valued measurable functions, so $\{s_n+h_n\}_{n\in N_+}$ is a sequence of simple measurable functions increasing to the non-negative valued function $f+g$. Finally, because measurability is preserved under limits, $f+g$ is measurable as well.

Similarly, because the product of simple functions is also a simple function and measurability is preserved over the product of real-valued measurable functions, $\{s_nh_n\}_{n\in N_+}$ is a sequence of simple measurable functions increasing to the non-negative function fg. Because measurability is preserved under limits, fg is measurable as well.

Q.E.D.

2.2.6 The Monotone Class Theorem for Functions

This section covers a very useful application of the $\pi-\lambda$ theorem.
Let $(E,\mathcal{E})$ be a measurable space. The colleciton $\mathcal{M}$ of $\mathcal{E}$-measurable numerical functions is said to be a monotone class if:

i) $I_E \in \mathcal{M}$

ii) For any bounded $f,g \in \mathcal{M}$ and $a,b \in \mathbb{R}$, $af+bg \in \mathbb{M}$

iii) For any increasing sequence $\{f_n\}_{n\in N_+}$ of non-negative functions in $\mathcal{M}$, the pointwise limit $f = \lim_{n\to\infty} f_n$ is contained in $\mathcal{M}$ (the limit exists because $\{f_n(x)\}_{n\in N_+}$ is increasing for any $x \in E$)

The following is the monotone class theorem:

Theorem 2.12 (The Monotone Class Theorem for Functions)
Let $(E,\mathcal{E})$ be a measurable space and $\mathcal{M}$ a monotone class of functions. If $\mathcal{M}$ contains every function of the form I_A, where $A \in \mathcal{E}_0$ and $\mathcal{E}_0$ is a π-system generating $\mathcal{E}$, then every non-negative or bounded $\mathcal{E}$-measurable function is contained in $\mathcal{M}$.

Proof) Define the collection

$$\mathcal{D} = \{A \subset E \mid I_A \in \mathcal{M}\}.$$

We will show that $\mathcal{D}$ is a λ-system on E:

i) $E \in \mathcal{D}$ because $I_E \in \mathcal{M}$ by the definition of a monotone class of functions.

ii) For any $A,B \in \mathcal{D}$ such that $A \subset B$, because $I_A, I_B \in \mathcal{M}$ are bounded and

$$I_{B\setminus A} = I_B - I_A,$$

the second property of monotone classes of functions tells us that $I_{B\setminus A} \in \mathcal{M}$ and $B \setminus A \in \mathcal{D}$.

iii) For any increasing sequence of sets $\{A_n\}_{n\in N_+} \subset \mathcal{D}$, letting $A = \bigcup_n A_n$, $I_{A_n} \le I_{A_{n+1}}$ for any $n \in N_+$ and

$$I_A(x) = \lim_{n\to\infty} I_{A_n}(x)$$

for any $x \in E$, so that $I_{A_n} \nearrow I_A$; because each $\{I_{A_n}\}_{n\in N_+}$ is an increasing sequence of non-negative functions in $\mathcal{M}$, $I_A \in \mathcal{M}$ by the third property of monotone classes of functions.

It is clear that $\mathcal{D}$ contains $\mathcal{E}_0$, since $I_A \in \mathcal{M}$ for any $A \in \mathcal{E}_0$ by hypothesis. $\mathcal{D}$ is a λ-system containing the π-system $\mathcal{E}_0$, so by the $\pi - \lambda$ theorem, $\mathcal{D}$ contains the σ-algebra generated by $\mathcal{E}_0$. Since $\mathcal{E}_0$ generates $\mathcal{E}$, $\mathcal{E} \subset \mathcal{D}$ and thus $I_A \in \mathcal{M}$ for any $A \in \mathcal{E}$.

The rest of the theorem will be proved by showing that the theorem holds for measurable simple functions and then, using theorem 2.10, for any non-negative measurable function. This type of argument will be employed heavily in the proofs to come.
Let $f \in \mathcal{E}_+$ be a measurable simple function given as $f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$ for $\alpha_1, \cdots, \alpha_n \in [0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{E}$. For any $1 \leq i \leq n$, $I_{A_i} \in \mathcal{M}$ is a bounded function, so by the second property of monotone classes,

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i} \in \mathcal{M}.$$

Now let $f \in \mathcal{E}_+$ in general. By theorem 2.10, there exists an increasing sequence $\{s_n\}_{n \in N_+}$ of $\mathcal{E}$-measurable simple functions such that $s_n \nearrow f$. Since each $s_n \in \mathcal{M}_+$, by the third property of monotone classes $f \in \mathcal{M}_+$ as well. Thus, $\mathcal{M}$ contains all non-negative $\mathcal{E}$-measurable functions.

Finally, let f be a bounded $\mathcal{E}$-measurable function. Then, $f^+, f^- \in \mathcal{E}_+$ by lemma 2.9, and because f^+, f^- are bounded,

$$f = f^+ - f^- \in \mathcal{M}$$

by the second property of monotone classes. Therefore, $\mathcal{M}$ also contains every bounded $\mathcal{E}$-measurable functions.
Q.E.D.

The monotone class theorem above will be used extensively in the context of product spaces and proofs involving measurable rectangles. By implication, it is also used extensively when dealing with transition kernels and conditional probabilities.

2.3 Measures

2.3.1 Definitions and Basic Properties

Let $(E, \mathcal{E})$ be a measurable space. Then, a measure μ on $\mathcal{E}$ is a function mapping $\mathcal{E}$ into $[0, +\infty]$ such that:

For any countable collection of disjoint measurable sets $\{A_n\}_{n \in N_+} \subset \mathcal{E}$,

$$\mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

The limit on the left is well-defined in $[0, +\infty]$ because $\{\mu(A_n)\}_{n \in N_+}$ is a non-negative sequence. This defining property of a measure is called countable additivity.
To avoid trivialities such as the fucntion $\mu(A) = +\infty$ for any $A \in \mathcal{E}$, we also require that

$$\mu(A) < +\infty \text{ for at least one } A \in \mathcal{E}.$$

The triple $(E, \mathcal{E}, \mu)$ is called a measure space.
The following properties follow immediately from the definition above:

Theorem 2.13 Let $(E, \mathcal{E}, \mu)$ be a measure space. Then, the following hold true:

i) $\mu(\emptyset) = 0$.

ii) (Finite Additivity) For any finite collection of disjoint measurable sets $A_1, \cdots, A_n \in \mathcal{E}$,

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

iii) (Monotonicity) For any $A, B \in \mathcal{E}$ such that $A \subset B$, $\mu(A) \leq \mu(B)$.

iv) (Sequential Continuity) For any sequence $\{A_n\}_{n \in N_+} \subset \mathcal{E}$ such that $A_n \subset A_{n+1}$ for any $n \in N_+$,

$$\mu\left(\bigcup_n A_n\right) = \lim_{n \to \infty} \mu(A_n).$$

v) (Sequential Continuity II) For any sequence $\{A_n\}_{n \in N_+} \subset \mathcal{E}$ such that $A_{n+1} \subset A_n$ for any $n \in N_+$ and $\mu(A_1) < +\infty$,

$$\mu\left(\bigcap_n A_n\right) = \lim_{n \to \infty} \mu(A_n).$$

vi) (Countable Subadditivity; Boole's Inequality) For any sequence $\{A_n\}_{n \in N_+} \subset \mathcal{E}$,

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

Proof) i) By definition, there exists a $A \in \mathcal{E}$ such that $\mu(A) < +\infty$. Define the sequence $\{A_n\}_{n\in N_+} \subset \mathcal{E}$ as $A_1 = A$ and $A_n = \emptyset$ for any $n > 1$. Then, $\{A_n\}_{n\in N_+}$ is disjoint, and by countable additivity,

$$\mu(A) = \mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \mu(A) + \sum_{n=2}^{\infty} \mu(\emptyset).$$

$\mu(A) < +\infty$, so $\sum_{n=2}^{\infty} \mu(\emptyset) = 0$. The only way this is possible is if $\mu(\emptyset) = 0$.

ii) Let $A_1, \cdots, A_n \in \mathcal{E}$ be a disjoint. Then, defining $A_m = \emptyset$ for any $m > n$, the sequence $\{A_m\}_{m\in N_+} \subset \mathcal{E}$ is disjoint, so by countable additivity,

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{n} A_i\right) &= \mu\left(\bigcup_m A_m\right) = \sum_{m=1}^{\infty} \mu(A_m) \\ &= \sum_{i=1}^{n} \mu(A_i) + \sum_{i=n+1}^{\infty} \mu(\emptyset) = \sum_{i=1}^{n} \mu(A_i)\end{aligned}$$

because $\mu(\emptyset) = 0$.

iii) Let $A, B \in \mathcal{E}$ and $A \subset B$. Then, $C = B \setminus A \in \mathcal{E}$ and $B = A \cup C$, where A and C are disjoint; by finite additivity,

$$\mu(B) = \mu(A \cup C) = \mu(A) + \mu(C) \geq \mu(A),$$

where the last inequality follows because $\mu(C) \in [0, +\infty]$.

iv) Let $\{A_n\}_{n\in N_+} \subset \mathcal{E}$ be an increasing sequence of sets. Define $\{B_n\}_{n\in N_+}$ as $B_1 = A_1$ and

$$B_n = A_n \setminus A_{n-1}$$

for any $n \geq 2$. Then, $\{B_n\}_{n\in N_+}$ is a sequence of disjoint measurable sets (due to the increasing nature of $\{A_n\}_{n\in N_+}$) such that $\bigcup_n B_n = \bigcup_n A_n = A \in \mathcal{E}$, and by countable additivity

$$\mu(A) = \mu\left(\bigcup_n B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i).$$

Recall that $B_1 = A_1$. Suppose, for some $n \geq 1$, that $A_m = \bigcup_{i=1}^{m} B_i$ for any $1 \leq m \leq n$. Then, because

$$A_{n+1} = B_{n+1} \cup A_n$$

and

$$A_n = \bigcup_{i=1}^{n} B_i$$

by the inductive hypothesis, we have

$$A_{n+1} = B_{n+1} \cup \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n+1} B_i.$$

Thus, by induction, $A_n = \bigcup_{i=1}^{n} B_i$ for any $n \in N_+$ and, by finite additivity,

$$\mu(A_n) = \sum_{i=1}^{n} \mu(B_i).$$

Therefore,

$$\mu(A) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \mu(A_n).$$

v) Let $\{A_n\}_{n\in N_+} \subset \mathcal{E}$ be a decreasing sequence of sets such that $\mu(A_1) < +\infty$. Define $\{B_n\}_{n\in N_+}$ as

$$B_n = A_1 \setminus A_n$$

for any $n \in N_+$; then, $\{B_n\}_{n\in N_+}$ is an increasing sequence of sets in $\mathcal{A}$ such that

$$\bigcup_n B_n = \bigcup_n (A_1 \setminus A_n) = \bigcup_n (A_1 \cap A_n^c) = A_1 \cap \left(\bigcup_n A_n^c \right) = A_1 \cap \left(\bigcap_n A_n \right)^c = A_1 \setminus \left(\bigcap_n A_n \right).$$

Denoting $A = (\bigcap_n A_n)$, by sequential continuity for increasing sets,

$$\mu(A_1 \setminus A) = \lim_{n\to\infty} \mu(B_n).$$

For any $n \in N_+$, $A_1 = A_n \cup B_n$, where A_n and B_n are disjoint, so by finite additivity, $\mu(A_1) = \mu(A_n) + \mu(B_n)$. By monotonicity and the fact that $A_n \subset A_1$, $\mu(A_n) \leq \mu(A_1) < +\infty$, which allows us to rewrite $\mu(B_n)$ as

$$\mu(B_n) = \mu(A_1) - \mu(A_n).$$

Likewise, because $A_1 = A \cup (A_1 \setminus A)$ for disjoint measurable sets A and $A_1 \setminus A$, finite additivity and monotonicity imply that

$$\mu(A) = \mu(A_1) - \mu(A_1 \setminus A).$$

Therefore,

$$\mu(A) = \mu(A_1) - \mu(A_1 \setminus A) = \mu(A_1) - \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \mu(A_n).$$

vi) Let $\{A_n\}_{n\in N_+} \subset \mathcal{E}$ be an arbitrary collection of measurable sets. Define $\{B_n\}_{n\in\mathcal{E}}$ as $B_1 = A_1$ and

$$B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right)$$

for any $n \geq 2$. Then, $\{B_n\}_{n\in N_+}$ is a sequence of disjoint measurable sets such that $\bigcup_n B_n = \bigcup_n A_n = A \in \mathcal{E}$, and by countable additivity

$$\mu(A) = \mu\left(\bigcup_n B_n\right) = \sum_{n=1}^{\infty} \mu(B_n).$$

Since $B_n \subset A_n$ for any $n \in N_+$, we have $\mu(B_n) \leq \mu(A_n)$ by monotonicity and thus

$$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Q.E.D.

2.3.2 Examples of Measures

The following are some measures that are often encountered:

(1) **The Lebesgue Measure**

The Lebesgue Measure λ on the euclidean space $\mathbb{R}^n$ is the measure such that, for any $A \in \mathcal{B}(\mathbb{R}^n)$, $\lambda(A)$ yields the length, area or volume of A. In particular, if $n=1$, $\lambda((a,b)) = b-a$ for any open interval (a,b); this also holds for closed or half-open intervals as well. Moreover, $\lambda(\{x\}) = 0$ for any $x \in \mathbb{R}^n$, that is, the Lebesgue measure assigns 0 mass to individual points in euclidean space.

The Lebesgue measure can be constructed directly (as in chapter 2 of the Frank Jones textbook), through the Caratheodory extension theorem (which we study below), or using the Riesz representation theorem (this is the avenue we pursue, and is detailed in chapter 4 of the present article).

(2) **The Counting Measure**

Let $(E, 2^E)$ be a countable space, where E is often taken to be N_+ or $\mathbb{Z}$. An important property of countable spaces is that, because the σ-algebra is often taken to be the discrete σ-algebra 2^E, every function defined on E is measurable.

The counting measure c on $(E, 2^E)$ is defined as

$$c(A) = \text{ The number of elements in } A$$

for any $A \subset E$. We can easily show that this is a measure:

i) $c(\emptyset) = 0$ because there are no elements in the empty set

ii) For any sequence $\{A_n\}_{n\in N_+}$ of disjoint subsets of E, let $A = \bigcup_n A_n$. If A is a finite set, then there exists an $N \in N_+$ such that $A_n = \emptyset$ for any $n \geq N$, and the number of elements in A is the sum of the number of elements in $A_1, \cdots, A_N$, since they are disjoint. This means that

$$c(A) = \sum_{n=1}^{N} c(A_n) = \sum_{n=1}^{\infty} c(A_n).$$

On the other hand, if A is an infinite set, we can consider two cases:
If there exists an $N \in N_+$ such that A_N is an infinite set, then because

$$c(A) = +\infty = \sum_{n=1}^{\infty} c(A_n).$$

If every set in $\{A_n\}_{n\in N_+}$ is a finite set, then only a finite number of sets in $\{A_n\}_{n\in N_+}$ is the empty set; otherwise, if there existed an $N \in N_+$ such that $A_n = \emptyset$ for any $n \geq N$, then A would contain only the finite number of elements contained in $A_1, \cdots, A_N$, which contradicts the assumption that A is an infinite set. Therefore, an infinite number of elements in the sequence $\{c(A_n)\}_{n\in N_+}$ is nonzero, which means that

$$\sum_{n=1}^{\infty} c(A_n) = +\infty = c(A).$$

In any case,

$$c(A) = \sum_{n=1}^{\infty} c(A_n),$$

so c is countably additive.

The counting measure will be used to great effect later on when interpreting series as integrals.

(3) **The Dirac Measure**

Let $(E,\mathcal{E})$ be a measurable space. For any $x \in E$, the Dirac Delta measure δ_x on $(E,\mathcal{E})$ sitting at x is defined as

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

for any $A \in \mathcal{E}$. To see that this actually defines a measure, note that

i) $\delta_x(\emptyset) = 0$ because $x \notin \emptyset$.

ii) For any sequence $\{A_n\}_{n\in N_+}$ of disjoint subsets of E, let $A = \bigcup_n A_n$. If $x \in A$, then there exists a unique $N \in N_+$ such that $x \in A_N$, where the uniqueness follows from the disjointness of $\{A_n\}_{n\in N_+}$; thus, $\delta_x(A_n) = 0$ for any $n \neq N$ because $x \notin A_n$, and

$$\delta_x(A) = 1 = \delta(A_N) = \sum_{n=1}^{\infty} \delta_x(A_n).$$

On the other hand, if $x \notin A$, then $x \notin A_n$ for every $n \in N_+$, and

$$\delta_x(A) = 0 = \sum_{n=1}^{\infty} \delta_x(A_n).$$

In any case,

$$\delta_x(A) = \sum_{n=1}^{\infty} \delta_x(A_n),$$

so δ_x is countably additive.

The Diract Delta measure has the tendency to pop up when we least expect it; examples include the proof of the existence of conditional probabilities and the transition probability of the Metropolis-Hastings algorithm.

(4) **Probability Measures**

Let $(\Omega, \mathcal{H})$ be a measurable space. A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{H})$ is a measure on $(\Omega, \mathcal{H})$ such that $\mathbb{P}(\Omega) = 1$, that is, the entire set is assigned the value of 1. In this case, the triple $(\Omega, \mathcal{H}, \mathbb{P})$ is said to be a probability space, where Ω is interpreted as the sample space (the set of all outcomes, such as heads or tails), $\mathcal{H}$ as the set of all admissible events, and $\mathbb{P}(H)$ as the probability of event H for any $H \in \mathcal{H}$.

2.3.3 The Finiteness of Measures

Note initially that the sum of measures is also a measure.
Let $\{\mu_n\}_{n\in N_+}$ be a sequence of measures on $(E,\mathcal{E})$. Define $\mu : \mathcal{E} \to [0,+\infty]$ as

$$\mu(A) = \sum_{n=1}^{\infty} \mu_n(A)$$

for any $A \in \mathcal{E}$, where the sum on the right hand side is well defined due to the non-negativity of each $\mu_n(A)$. We will show that μ is actually a measure on $(E,\mathcal{E})$:

- $\mu(\emptyset) = \sum_{n=1}^{\infty} \mu_n(\emptyset) = 0$ because each $\mu_n(\emptyset) = 0$. As such, $\mu(\emptyset) < +\infty$ for $\emptyset \in \mathcal{E}$.
- For any disjoint $\{A_m\}_{m\in N_+} \subset \mathcal{E}$,

$$\begin{aligned}
\mu\left(\bigcup_m A_m\right) &= \sum_{n=1}^{\infty} \mu_n\left(\bigcup_m A_m\right) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu_n(A_m) && \text{(Countable Additivity)} \\
&= \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \mu_n(A_m) \\
&\quad \text{(The order of summation of non-negative sequences can be changed)} \\
&= \sum_{m=1}^{\infty} \mu(A_m). && \text{(By definition)}
\end{aligned}$$

Therefore, μ is a measure on $(E,\mathcal{E})$.

Since the function $\nu : \mathcal{E} \to [0,+\infty]$ defined as $\nu(A) = 0$ for any $A \in \mathcal{E}$ is trivially a measure on $(E,\mathcal{E})$, it follows from logic similar to that used to prove the finite additivity of measures that the sum of finite measures is also a measure.

The above fact is used to prove the relationship between different kinds of measures.
Let μ be a measure on $(E,\mathcal{E})$. Then, we say that μ is:

- finite if $\mu(E) < +\infty$,
- σ-finite if there exists a measurable and countable partition $\{E_n\}_{n\in N_+} \subset \mathcal{E}$ of E such that $\mu(E_n) < +\infty$ for any $n \in N_+$.
- Σ-finite if there exists a sequence of finite measures $\{\mu_n\}_{n\in N_+}$ on $(E,\mathcal{E})$ such that

$$\mu = \sum_{n=1}^{\infty} \mu_n.$$

An example of a finite measure is a probability measure, which must satisfy $\mu(E) = 1 < +\infty$ by definition.
An example of a σ-fintie measure is the Lebesgue measure λ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$; note that the collection $\{B(\mathbf{0}, n+1) \setminus B(\mathbf{0}, n)\}_{n\in N_+}$, where $B(x,\delta)$ is the open ball of radius $\delta > 0$ around $x \in \mathbb{R}^n$,

is a countable and measurable partition of $\mathbb{R}^n$ and that the Lebesgue measure of each entry is finite (the entries have finite volume).
The counting measure is also a σ-finite measure.

We can show that finite measures are σ-finite, and that σ-finite measures are Σ-finite.

Lemma 2.14 Let $(E, \mathcal{E}, \mu)$ be a measure space. Then, the following hold true:

i) If μ is finite, then it is σ-finite.

ii) If μ is σ-finite, then it is Σ-finite.

Proof) i) Suppose that $\mu(E) < +\infty$. Then, $\{E_n\}_{n\in N_+}$ defined as $E_1 = E$ and $E_n = \emptyset$ for $n \geq 2$ is a measurable partition of E such that $\mu(E_n) < +\infty$ for any $n \in N_+$, so by definition μ is σ-finite.

ii) Suppose that μ is σ-finite, so that there exists a measurable partition $\{E_n\}_{n\in N_+} \subset \mathcal{E}$ of E such that $\mu(E_n) < +\infty$ for any $n \in N_+$. For any $n \in N_+$, define $\mu_n : \mathcal{E} \to [0, +\infty]$ as

$$\mu_n(A) = \mu(A \cap E_n)$$

for any $A \in \mathcal{E}$. Then, μ_n is a finite measure on $(E, \mathcal{E})$:

- $\mu_n(\emptyset) = \mu(\emptyset \cap E_n) = \mu(\emptyset) = 0$,
- For any disjoint $\{A_m\}_{m\in N_+} \subset \mathcal{E}$,

$$\begin{aligned}
\mu_n\left(\bigcup_m A_m\right) &= \mu\left(\left(\bigcup_m A_m\right) \cap E_n\right) = \mu\left(\bigcup_m (A_m \cap E_n)\right) \\
&= \sum_{m=1}^{\infty} \mu(A_m \cap E_n) \\
&\qquad \text{(Countable additivity; } \{A_m \cap E_n\}_{m\in N_+} \text{ is a disjoint collection)} \\
&= \sum_{m=1}^{\infty} \mu_n(A_m),
\end{aligned}$$

- $\mu_n(E) = \mu(E_n) < +\infty$.

Now observe that, for any $A \in N_+$,

$$\begin{aligned}
\mu(A) &= \mu\left(A \cap \left(\bigcup_n E_n\right)\right) && (\textstyle\bigcup_n E_n = E) \\
&= \mu\left(\bigcup_n (A \cap E_n)\right) \\
&= \sum_{n=1}^{\infty} \mu(A \cap E_n) \\
& \text{(Countable additivity; } \{A \cap E_n\}_{m \in N_+} \text{ is a disjoint collection)} \\
&= \sum_{n=1}^{\infty} \mu_n(A).
\end{aligned}$$

As such, $\mu = \sum_{n=1}^{\infty} \mu_n$, and μ is Σ-finite by definition.

Q.E.D.

The $\pi - \lambda$ theorem allows us to prove the following result, which enables us to determine whether two finite or σ-finite measures are equal on the basis of their equality on a generating set of the relevant σ-algebra.

Lemma 2.15 Let $(E, \mathcal{E})$ be a measurable space, and μ, ν two measures on $(E, \mathcal{E})$. Let $\mathcal{F}$ be a π-system that generates $\mathcal{E}$. Then, the following hold true:

i) Suppose μ, ν are finite measures.
If $\mu(A) = \nu(A)$ for any $A \in \mathcal{F}$ and $\mu(E) = \nu(E) < +\infty$, then $\mu(A) = \nu(A)$ for any $A \in \mathcal{E}$.

ii) Now suppose μ, ν are σ-finite measures.
If $\mu(A) = \nu(A)$ for any $A \in \mathcal{F}$ and $\mathcal{F}$ contains a measurable partition $\{E_n\}_{n \in N_+}$ of E such that $\mu(E_n) = \nu(E_n) < +\infty$ for any $n \in N_+$, then $\mu(A) = \nu(A)$ for any $A \in \mathcal{E}$.

Proof) i) Define the collection

$$\mathcal{M} = \{A \subset E \mid \mu(A) = \nu(A)\}.$$

We will show that $\mathcal{M}$ is a λ-system:

- $E \in \mathcal{M}$ because $\mu(E) = \nu(E) < +\infty$.

- For any $A, B \in \mathcal{M}$ such that $A \subset B$, by finite additivity and the fact that $\mu(A) = \nu(A) < +\infty$ and $\mu(B) = \nu(B) < +\infty$, we have

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A).$$

Therefore, $B \setminus A \in \mathcal{M}$.

– For any increasing sequence of sets $\{A_n\}_{n\in N_+} \subset \mathcal{M}$, letting $A = \bigcup_n A_n$, by sequential continuity

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(A).$$

Therefore, $A \in \mathcal{M}$.

The π-system $\mathcal{F}$ is contained in the λ-system $\mathcal{M}$; by the $\pi - \lambda$ theorem, it follows that the σ-algebra $\mathcal{E}$ generated by $\mathcal{F}$ is contained in $\mathcal{M}$, so that $\mu(A) = \nu(A)$ for any $A \in \mathcal{E}$.

ii) For any $n \in N_+$, define $\mu_n, \nu_n : E \to [0, +\infty]$ as

$$\mu_n(A) = \mu(A \cap E_n) \quad \text{and} \quad \nu_n(A) = \nu(A \cap E_n)$$

for any $n \in N_+$. By assumption, $\{E_n\}_{n\in N_+}$ is a measurable partition of E such that $\mu(E_n) = \nu(E_n) < +\infty$; we showed in the proof of lemma 2.14 that each μ_n, ν_n are finite measures on $(E, \mathcal{E})$ such that

$$\mu = \sum_{n=1}^{\infty} \mu_n \quad \text{and} \quad \nu = \sum_{n=1}^{\infty} \nu_n.$$

For any $n \in N_+$, μ_n and ν_n are finite measures such that $\mu_n(E) = \mu(E_n) = \nu(E_n) = \nu_n(E) < +\infty$ and, because E_n is an element of the π-system $\mathcal{F}$,

$$\mu_n(A) = \mu(A \cap E_n) = \nu(A \cap E_n) = \nu_n(A)$$

for any $A \in \mathcal{F}$. By result *i)*, $\mu_n(A) = \nu_n(A)$ for any $A \in \mathcal{E}$. It now follows that, for any $A \in \mathcal{E}$,

$$\mu(A) = \sum_{n=1}^{\infty} \mu_n(A) = \sum_{n=1}^{\infty} \nu_n(A) = \nu(A).$$

Q.E.D.

2.3.4 Caratheodory's Extension Theorem

We have studied the definition and properties of measures, but a central question we have not answered yet is whether measures actually exist. Given a measurable space $(E, \mathcal{E})$, is there really a non-negative function defined on $\mathcal{E}$ that is countably additive? This is the question we seeek to answer in this section. Looking ahead slightly, it turns out that, given pre-measures, or functions that are countably additive on some algebra, we can extend that algebra to a σ-algebra and the pre-measure to a measure. This result, called Caratheodory's extension theorem, is the focus of our section.
The existence of pre-measures is easily established for a small algebra of sets, so the existence of a measure on an extension of such algebras to σ-algebras follows from the extension theorem. In practice, this theorem is used in a variety of contexts in probability theory, especially in the proof of the existence of chains of random variables and Prohorov's theorem, which furnishes sufficient conditions for the weak convergence of probability measures on a metric space.

Recall that an algebra $\mathcal{A}$ on a set E is a collection of subsets of E satisfying the following properties:

- $E \in \mathcal{A}$.
- For any $A \in \mathcal{A}$, $A^c \in \mathcal{A}$. This implies that $\emptyset \in \mathcal{A}$ as well.
- For any finite collection $\{A_1, \cdots, A_n\} \subset \mathcal{A}$, their union $A = \bigcup_{i=1}^n A_i \in \mathcal{A}$ as well. By property ii), $\mathcal{A}$ is also closed under finite intersections.

A pre-measure μ_0 on $\mathcal{A}$ is defined as a function $\mu_0 : \mathcal{A} \to [0, +\infty]$ such that

- $\mu_0(\emptyset) = 0$.
- **(Finite Additivity)** For any finite disjoint set $\{A_1, \cdots, A_n\} \subset \mathcal{A}$,

$$\mu_0\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu_0(A_i).$$

Any pre-measure μ_0 on $\mathcal{A}$ is said to have the σ-additivity property if, for any sequence $\{A_n\}_{n \in N_+}$ of disjoint sets in $\mathcal{A}$ with union A also in $\mathcal{A}$,

$$\mu_0(A) = \sum_{n=1}^\infty \mu_0(A_n).$$

Caratheodory's extension theorem shows that, given an algebra $\mathcal{A}$ on E and a σ-additive pre-measure μ_0 on $\mathcal{A}$, $\mathcal{A}$ can be extended to a σ-algebra on E and μ_0 to a measure on that σ-algebra. The exact sense in which these are "extensions" will soon be made clear.

An outer measure on E is defined as a function $\mu : 2^E \to [0, +\infty]$ on 2^E such that:

- $\mu(\emptyset) = 0$,
- **(Monotonicity)** For any subsets A, B of E such that $A \subset B$, we have $\mu(A) \leq \mu(B)$.
- **(σ-Subadditivity)** For any sequence $\{A_n\}_{n \in N_+}$ of subsets of E with union A,

$$\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Let μ^* be an outer measure on E. The collection of all μ^*-measurable sets is defined as

$$\mathcal{M} = \{A \subset E \mid \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) \text{ for any } B \subset E\}.$$

In a sense, $\mathcal{M}$ collects all the subsets of E that partitions each subset of E in a manner that is finitely additive under μ^*.
Our main result, to be proven immediately below, states that the μ^*-measurable sets form a σ-algebra on E, and that the restriction of μ^* to this σ-algebra is a measure. From this result, it follows that, if we can just construct an outer measure from a pre-measure on some algebra, then we have the desired construction.

Theorem 2.16 (Caratheodory's Restriction Theorem)
Let E be an arbitrary set, μ^* an outer measure on E, and $\mathcal{M}$ the collection of all μ^*-measurable sets. Then, $\mathcal{M}$ is a σ-algebra on E, and the restriction μ of μ^* to $\mathcal{M}$ is a measure on $\mathcal{M}$. In addition, the measure space $(E, \mathcal{M}, \mu)$ is complete.

Proof) We will prove the two results above simultaneously.
Clearly, $E \in \mathcal{M}$, since

$$\mu^*(B) = \mu^*(B \cap E) = \mu^*(B \cap E) + \mu^*(B \cap \emptyset)$$

for any $B \subset E$.
Next, we want to show that $\mathcal{M}$ is closed under complements. Suppose $A \in \mathcal{M}$. Then,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

for any $B \subset E$, which immediately shows us that $A^c \in \mathcal{M}$. It follows that $\emptyset \in \mathcal{M}$ as well.

Closedness under Finite Unions and Finite Additivity
To show that $\mathcal{M}$ is closed under countable unions, we first show that it is closed under

finite unions.
Let $A_1, A_2 \in \mathcal{M}$, and let $A = A_1 \cup A_2$. Then, for any $B \subset E$,

$$\mu^*(B) = \mu^*((B\cap A)\cup(B\cap A^c)) \leq \mu^*(B\cap A)+\mu^*(B\cap A^c).$$

by the subadditivity of μ^*. Therefore, we need only show the reverse inequality holds to conclude that $A \in \mathcal{M}$. This follows easily:

$$\begin{aligned}
\mu^*(B) &= \mu^*(B\cap A_1)+\mu^*(B\cap A_1^c) && (A_1 \in \mathcal{M})\\
&= \mu^*(B\cap A_1)+\mu^*(B\cap A_1^c\cap A_2)+\mu^*(B\cap A_1^c\cap A_2^c) && (B\cap A_1^c \subset E \text{ and } A_2 \in \mathcal{M})\\
&\geq \mu^*((B\cap A_1)\cup(B\cap A_1^c\cap A_2))+\mu^*(B\cap A_1^c\cap A_2^c) && (\text{Subadditivity of } \mu*)\\
&= \mu^*(B\cap(A_1\cup(A_1^c\cap A_2)))+\mu^*(B\cap A_1^c\cap A_2^c)\\
&= \mu^*(B\cap(A_1\cup A_2))+\mu^*(B\cap A_1^c\cap A_2^c)\\
&= \mu^*(B\cap A)+\mu^*(B\cap A^c).
\end{aligned}$$

Therefore, $A \in \mathcal{M}$.
Furthermore, if A_1, A_2 are disjoint, then taking $A = A_1 \cup A_2$ again,

$$\begin{aligned}
\mu^*(B\cap A) &= \mu^*((B\cap A)\cap A_1)+\mu^*((B\cap A)\cap A_1^c)\\
&= \mu^*(B\cap A_1\cup((B\cap A_2)\cap A_1))+\mu^*((B\cap A_2)\cap A_1^c)\\
&= \mu^*(B\cap A_1)+\mu^*(B\cap A_2).
\end{aligned}$$

for any $B \subset E$. Taking $B = E$ show us that μ^* is finitely additive on $\mathcal{M}$.
These results can easily be extended to finite collections of sets in $\mathcal{M}$ through induction; specifically, for any $\{A_1, \cdots, A_n\} \subset \mathcal{M}$,

$$\bigcup_{i=1}^{n} A_i \in \mathcal{M}$$

$$\mu^*\left(B\cap\left(\bigcup_{i=1}^{n} A_i\right)\right) = \sum_{i=1}^{n}\mu^*(B\cap A_i). \qquad (\text{for any } B \subset E \text{ if } A_1, \cdots, A_n \text{ are disjoint})$$

A consequence of the closedness of $\mathcal{M}$ under finite unions and complements is that it is also closed under set differences. To see this, let $A_1, A_2 \in \mathcal{M}$. Then, $A_1 \setminus A_2 = A_1 \cap A_2^c = (A_1^c \cup A_2)^c$; since $A_1^c \in \mathcal{M}$ and $A_1^c \cup A_2 \in \mathcal{M}$, it follows that $A_1 \setminus A_2 \in \mathcal{M}$ as well.

Closedness under Countable Unions and Countable Additivity

Now let $\{A_n\}_{n\in N_+}$ be a sequence of sets in $\mathcal{M}$ with union A. Again, by the subadditivty of μ^*,

$$\mu^*(B) \leq \mu^*(B\cap A)+\mu^*(B\cap A^c)$$

for any $B \subset E$. For the reverse inequality, define the sequence $\{V_n\}_{n\in N_+}$ as $V_1 = A_1$ and

$$V_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right)$$

for any $n \geq 2$. Because $\mathcal{M}$ is closed under finite unions and set differences, $\{V_n\}_{n\in N_+}$ is a disjoint sequence of sets in $\mathcal{M}$ such that

$$\bigcup_{i=1}^{n} V_i \nearrow A$$

as $n \to \infty$. Therefore, for any $n \in N_+$,

$$\begin{aligned}
\mu^*(B) &= \mu^*\left(B \cap \left(\bigcup_{i=1}^{n} V_i\right)\right) + \mu^*\left(B \cap \left(\bigcup_{i=1}^{n} V_i\right)^c\right) \\
&\geq \mu^*\left(B \cap \left(\bigcup_{i=1}^{n} V_i\right)\right) + \mu^*(B \cap A^c) \\
&= \sum_{i=1}^{n} \mu^*(B \cap V_i) + \mu^*(B \cap A^c). \qquad (V_1, \cdots, V_n \text{ are disjoint sets in } \mathcal{M})
\end{aligned}$$

This holds for any $n \in N_+$, so taking $n \to \infty$ on both sides,

$$\mu^*(B) \geq \sum_{n=1}^{\infty} \mu^*(B \cap V_n) + \mu^*(B \cap A^c),$$

and by the countable subadditivity of μ^*,

$$\begin{aligned}
\mu^*(B) &\geq \sum_{n=1}^{\infty} \mu^*(B \cap V_n) + \mu^*(B \cap A^c) \\
&\geq \mu^*\left(\bigcup_{n} (B \cap V_n)\right) + \mu^*(B \cap A^c) \\
&= \mu^*\left(B \cap \left(\bigcup_{n} V_n\right)\right) + \mu^*(B \cap A^c) = \mu^*(B \cap A) + \mu^*(B \cap A^c).
\end{aligned}$$

This holds for any $B \subset E$, so it follows that $A \in \mathcal{M}$.
Finally, taking $B = A$ reveals that

$$\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A \cap V_n) + \mu^*(A \cap A^c) = \sum_{n=1}^{\infty} \mu^*(V_n).$$

If $\{A_n\}_{n\in N_+}$ were disjoint, then $V_n = A_n$ for any $n \in N_+$, so in this case, we have

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

and the reverse inequality holds by countable subadditivity. Therefore,

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

if $\{A_n\}_{n\in N_+}$ is disjoint, which tells us that μ^* is countably additive on $\mathcal{M}$.

Completeness

We have thus shown that $(E, \mathcal{M})$ is a measurable space, and letting $\mu : \mathcal{M} \to [0, +\infty]$ be the restriction of μ^* to $\mathcal{M}$, because $\mu(\emptyset) = \mu^*(\emptyset) = 0$ and μ is countably additive on $\mathcal{M}$, the triple $(E, \mathcal{M}, \mu)$ defines a measure space.

It remains to show that this measure space is complete. Choose any $A \in \mathcal{M}$ such that $\mu(A) = 0$, and let $N \subset A$. By the monotonicity of μ^*, $\mu^*(N) = 0$, which implies that

$$\mu^*(B \cup N) \leq \mu^*(B) + \mu^*(N) = \mu^*(B) \leq \mu^*(B \cup N)$$

and therefore $\mu^*(B) = \mu^*(B \cup N)$ for any $B \subset E$ by the monotonicity and subadditivity of μ^*. By subadditivity,

$$\mu^*(B) = \mu^*(B \cup N) = \mu^*((B \setminus N) \cup N) \leq \mu^*(B \setminus N) + \mu^*(N) = \mu^*(B \cap N^c),$$

and because $B \cap N^c \subset B$, the reverse inequality holds as well, which implies that

$$\mu^*(B) = \mu^*(B \cap N^c).$$

$\mu^*(B \cap N) = 0$ trivially, so it follows that

$$\mu^*(B) = \mu^*(B \cap N) + \mu^*(B \cap N^c).$$

Therefore, $N \in \mathcal{M}$ by definition, which implies that

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

It follows that $\mathcal{M}$ contains all negligible sets, and that their measure is 0 under μ; by definition, $(E, \mathcal{M}, \mu)$ is complete.

Q.E.D.

Now we show a way to define an outer measure for any set E given an algebra on E and a pre-measure on that algebra.

Lemma 2.17 Let E be an arbitrary set, $\mathcal{A}$ an algebra on E, and μ_0 a pre-meausure on $\mathcal{A}$.

Then, the function $\mu^* : 2^E \to [0,+\infty]$ defined as

$$\mu^*(A) = \inf\Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } A \subset \bigcup_n V_n\Big\}$$

for any $A \subset E$ is an outer measure on E.

Proof) $\mu^*(A)$ is clearly well-defined for any $A \subset E$; $\{V_n\}_{n\in N_+}$ where $V_1 = E$ and $V_n = \emptyset$ for any $n \geq 2$ is a collection of elements of $\mathcal{A}$ such that $A \subset \bigcup_n V_n$ and

$$\sum_{n=1}^{\infty}\mu_0(V_n) = \mu_0(E) \in [0,+\infty].$$

Therefore, the set

$$\Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } A \subset \bigcup_n V_n\Big\}$$

contains $\mu_0(E)$, and is thus a nonempty subset of $[0,+\infty]$. If the set contains a finite element, then by the lower bound property of the real line, $\mu^*(A) \in [0,+\infty)$, while if the set equals $\{+\infty\}$, then $\mu^*(A) = +\infty$.

Now we show that μ^* satisfies the three properties of an outer measure. Because $\{V_n\}_{n\in N_+}$ defined as $V_n = \emptyset$ for any $n \in N_+$ is a sequence in $\mathcal{A}$ that covers $\emptyset$, it follows that

$$0 \leq \mu^*(\emptyset) \leq 0,$$

and as such $\mu^*(\emptyset) = 0$.
Next, let A, B be subsets of E such that $A \subset B$. For any collection $\{V_n\}_{n\in N_+} \subset \mathcal{A}$ such that $B \subset \bigcup_n V_n$, it follows that $A \subset \bigcup_n V_n$ as well, so that

$$\Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } B \subset \bigcup_n V_n\Big\} \subset \Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } A \subset \bigcup_n V_n\Big\}.$$

Therefore,

$$\begin{aligned}\mu^*(A) &= \inf\Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } A \subset \bigcup_n V_n\Big\} \\ &\leq \inf\Big\{\sum_{n=1}^{\infty}\mu_0(V_n) \mid \{V_n\}_{n\in N_+} \subset \mathcal{A} \text{ and } B \subset \bigcup_n V_n\Big\} = \mu^*(B),\end{aligned}$$

which shows that μ^* is monotonic.

It remains to show that μ^* is countably subadditive. To this end, choose any sequence

$\{A_n\}_{n\in N_+}$ of subsets of E with union A. If $\mu^*(A_n) = +\infty$ for some $n \in N_+$, then the reuslt is trivial.
Assume then that $\mu^*(A_n) < +\infty$ for any $n \in N_+$. For any $\varepsilon > 0$ and $n \in N_+$, by the definition of the infimum there exists a sequence $\{V_{m,n}\}_{m\in N_+} \subset \mathcal{A}$ such that $A_n \subset \bigcup_m V_{m,n}$ and

$$\mu^*(A_n) \le \sum_{m=1}^{\infty} \mu_0(V_{m,n}) < \mu^*(A_n) + \frac{\varepsilon}{2^n}.$$

Then, the sequence $\{V_{m,n}\}_{m,n\in N_+}$ is a sequence of sets in $\mathcal{A}$ that covers A, since

$$A = \bigcup_n A_n \subset \bigcup_n \bigcup_m V_{m,n}.$$

Furthermore, because $\mu_0(V_{m,n})$ is non-negative for any $m,n \in N_+$, the order of summation can be freely interchanged in the series $\sum\sum \mu_0(V_{m,n})$, which tells us, by the definition of μ^*,

$$\begin{aligned}\mu^*(A) \le \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu_0(V_{m,n}) &\le \sum_{n=1}^{\infty} \mu^*(A_n) + \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} \\ &= \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.\end{aligned}$$

This holds for any $\varepsilon > 0$, so

$$\mu^*(A) \le \sum_{n=1}^{\infty} \mu^*(A_n).$$

This completes the proof.
Q.E.D.

We can now state the final extension result:

Theorem 2.18 (Hahn-Kolmogorov Extension Theorem)
Let E be an arbitrary set, $\mathcal{A}$ an algebra on E, and μ_0 a σ-additive pre-meausure on $\mathcal{A}$. Then, there exists a σ-algebra $\mathcal{M}$ on E and a measure μ on $(E,\mathcal{M})$ such that:

i) $(E,\mathcal{M},\mu)$ is a complete measure space.

ii) $\mathcal{M}$ contains $\mathcal{A}$.

iii) $\mu(A) = \mu_0(A)$ for any $A \in \mathcal{A}$.

Proof) Let μ^* be the outer measure on E defined as in lemma 2.17, and let $\mathcal{M}$ be the collection of all μ^*-measurable sets and μ the restriction of μ^* to $\mathcal{M}$. Then, by Caratheodory's restriction theorem, $(E,\mathcal{M},\mu)$ is a complete measure space. It remains to verify the last two claims, which contextualize why $(E, \mathcal{M}, \mu)$ is an "extension" of the algebra $\mathcal{A}$ and the pre-measure μ_0.

We first show that $\mathcal{A} \subset \mathcal{M}$.
Choose any $B \subset E$ and $A \in \mathcal{A}$. Then,

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

by the subadditivty of μ^*, so in order to show that $A \in \mathcal{M}$, we must show that the reverse inequality holds as well.
If $\mu^*(B) = +\infty$, then the reverse inequality holds trivially. Suppose $\mu^*(B) < +\infty$; then, for any $\varepsilon > 0$ there exists an $\mathcal{A}$-cover $\{V_n\}_{n \in N_+}$ of B such that

$$\mu^*(B) \leq \sum_{n=1}^{\infty} \mu_0(V_n) < \mu^*(B) + \varepsilon.$$

It follows that $\{A \cap V_n\}_{n \in N_+}$ is an $\mathcal{A}$-cover of $B \cap A$, so that

$$\mu^*(B \cap A) \leq \sum_{n=1}^{\infty} \mu_0(A \cap V_n);$$

likewise,

$$\mu^*(B \cap A^c) \leq \sum_{n=1}^{\infty} \mu_0(A^c \cap V_n).$$

Adding the two equations together yields

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum_{n=1}^{\infty} \mu_0(A \cap V_n) + \sum_{n=1}^{\infty} \mu_0(A^c \cap V_n) \\ &= \sum_{n=1}^{\infty} \left[\mu_0(A \cap V_n) + \mu_0(A^c \cap V_n)\right], \end{aligned}$$

where the equality follows because the series in question both converge. Because $A \cap V_n$, $A^c \cap V_n$ are disjoint sets in $\mathcal{A}$, by the finite additivity of μ_0

$$\mu_0(A \cap V_n) + \mu_0(A^c \cap V_n) = \mu_0(V_n),$$

and as such

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \sum_{n=1}^{\infty} \mu_0(V_n) < \mu^*(B) + \varepsilon.$$

This holds for any $\varepsilon > 0$, so

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \mu^*(B).$$

This shows us that $A \in \mathcal{M}$, so the algebra $\mathcal{A}$ is contained in $\mathcal{M}$.

Next, we show that μ^* and μ_0 agree on $\mathcal{A}$ due to the σ-additivity of μ_0.
Choose any $A \in \mathcal{A}$. Then, because the sequence $\{V_n\}_{n \in N_+}$ defined as $V_1 = A$ and $V_n = \emptyset$ for any $n \geq 2$ is a $\mathcal{A}$-cover of A, by the definition of μ^* we have

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \mu_0(V_n) = \mu_0(A).$$

Now consider an arbitrary $\mathcal{A}$-cover $\{V_n\}_{n \in N_+}$ of A. By definition, $A \subset \bigcup_n V_n$, so

$$\bigcup_n (A \cap V_n) = A.$$

Define the collection $\{B_n\}_{n \in N_+}$ as follows:

$$\begin{aligned} B_1 &= A \cap V_1 \\ B_n &= (A \cap V_n) \setminus \left(\bigcup_{i=1}^{n-1} (A \cap V_i) \right) \quad \text{for any } n \geq 2. \end{aligned}$$

Since algebras are closed under finite intersections, $\{A \cap V_n\}_{n \in N_+}$ is a sequence in $\mathcal{A}$. Furthermore, they are also closed under finite unions and set differences, so $\{B_n\}_{n \in N_+}$ is a disjoint sequence of sets in $\mathcal{A}$ with union A. By the σ-additivity of μ_0,

$$\mu_0(A) = \sum_{n=1}^{\infty} \mu_0(B_n),$$

and because $B_n \subset A \cap V_n \subset V_n$ for any $n \in N_+$, the monotonicity of the pre-measure μ_0 tells us that

$$\mu_0(A) \leq \sum_{n=1}^{\infty} \mu_0(V_n).$$

Therefore, $\mu_0(A)$ is a lower bound of the set

$$\left\{ \sum_{n=1}^{\infty} \mu_0(V_n) \mid \{V_n\}_{n \in N_+} \subset \mathcal{A} \text{ and } A \subset \bigcup_n V_n \right\},$$

which impliest that $\mu_0(A) \leq \mu^*(A)$. It follows that

$$\mu^*(A) = \mu_0(A).$$

Because $A \in \mathcal{A} \subset \mathcal{M}$, we finally have the equality

$$\mu_0(A) = \mu(A).$$

Q.E.D.

Note that, in the above proof, the σ-additivity of μ_0 was only used to show that μ and μ_0 agree on $\mathcal{A}$.
Therefore, we can conclude that, for any algebra $\mathcal{A}$ on E and a pre-measure μ_0 on $\mathcal{A}$, there exists a complete measure space $(E, \mathcal{M}, \mu)$ such that $\mathcal{A} \subset \mathcal{M}$ and $\mu(A) \leq \mu_0(A)$ for any $A \in \mathcal{A}$.

The Hahn-Kolmorov Extension theorem tells us only that there exists a complete measure space that extends the given algebra $\mathcal{A}$ and pre-measure μ_0 in the manners specified above. These are quite weak conditions, and do not impart much information about the constructed measure space. This is sometimes an advantage, especially in probability theory, where we want the constructed probability space to be as general as possible.
In other cases, however, the paucity of information proves detrimental. In particular, while it is possible to construct the Lebesgue measure using the extension theorem above, doing so means that we have to prove its other properties, such as regularity and the approximation property, as well as the equivalence of Lebesgue integrals and Riemann integrals, separately. To this end, we study the Riesz representation theorem in chapter 4. This theorem allows us to extend a linear functional defined on the space of continuous compactly supported functions to a measure, as the above theorem extended a pre-measure to a measure. In the case of the Riesz theorem, however, the constructed measure is always a Radon measure and satisfies several desirable properties off the bat, which makes it ideal to construct measures such as the Lebesgue measure.

Chapter 3

Abstract Integration

3.1 Integration of Non-Negative Functions

Let $(E, \mathcal{E}, \mu)$ be a measure space and $f \in \mathcal{E}_+$ a measurable simple function given as

$$f = \sum_{i=1}^{n} \alpha_i I_{A_i}$$

for some $\alpha_1, \cdots, \alpha_n \in [0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{E}$.
For any $A \in \mathcal{E}$, we define the integral of f with respect to μ over A as

$$\int_A f d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A);$$

note that this value takes values in $[0, +\infty]$.

Before we start using this definition in earnest, we must first verify some of its properties, the most important of which is whether it is invariant to the representation of the simple function as a linear combination of a finite number of indicator functions.

Lemma 3.1 Let $(E, \mathcal{E}, \mu)$ be a measure space and $f, g \in \mathcal{E}_+$ be measurable simple functions. The following hold true:

i) For any $A \in \mathcal{E}$, the mapping $f \to \int_A f d\mu$ is a function.

ii) (Monotonicity) If $f \leq g$, then for any $A \in \mathcal{E}$, $\int_A f d\mu \leq \int_A g d\mu$.

iii) (Linearity) For any $c \in [0, +\infty)$ and $A \in \mathcal{E}$, $\int_A (cf + g) d\mu = c \cdot \int_A f d\mu + \int_A g d\mu$.

Proof) We first prove $ii)$, and then show how it implies $i)$.
Suppose the measurable simple functions f, g satisfy $f \leq g$ and have the canonical representations

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i} \quad \text{and} \quad g = \sum_{i=1}^{m} \beta_i \cdot I_{B_i}.$$

As in lemma 2.11, define $C_{ij} = A_i \cap B_j$ for any $1 \le i \le n$, $1 \le j \le m$; note that the C_{ij} are all disjoint and cover E because $A_1, \cdots, A_n$ and $B_1, \cdots, B_m$ are disjoint and they each cover E. In other words, any $x \in E$ is contained in C_{ij} for exactly one pair (i,j). Choose any $A \in \mathcal{E}$, and partition the set $I = \{(i,j) \mid 1 \le i \le n, 1 \le j \le m\}$ into two subsets:

$$P_1 = \{(i,j) \in I \mid C_{ij} \cap A \neq \emptyset\} \quad \text{and} \quad P_2 = \{(i,j) \in I \mid C_{ij} \cap A = \emptyset\}.$$

For any $(i,j) \in P_1$, since $f(x) \le g(x)$ for any $x \in E$, for any $x \in C_{ij}$

$$f(x) = \alpha_i \le \beta_j = g(x).$$

Observe that, by finite additivity and the fact that $E = \bigcup_{j=1}^m B_j$,

$$\begin{aligned} \mu(A_i \cap A) &= \mu\left(A_i \cap \left(\bigcup_{j=1}^m B_j\right) \cap A\right) \\ &= \mu\left(\bigcup_{j=1}^m (C_{ij} \cap A)\right) \\ &= \sum_{j=1}^m \mu(C_{ij} \cap A), \end{aligned}$$

for any $1 \le i \le n$, so that

$$\int_A f d\mu = \sum_{i=1}^n \alpha_i \cdot \mu(A_i \cap A) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \cdot \mu(C_{ij} \cap A).$$

Likewise, we can see that

$$\int_A g d\mu = \sum_{j=1}^m \beta_j \cdot \mu(B_j \cap A) = \sum_{i=1}^n \sum_{j=1}^m \beta_i \cdot \mu(C_{ij} \cap A).$$

Therefore,

$$\begin{aligned} \int_A f d\mu &= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \cdot \mu(C_{ij} \cap A) \\ &= \sum_{(i,j) \in P_1} \alpha_i \cdot \mu(C_{ij} \cap A) && (\mu(C_{ij} \cap A) = 0 \text{ for any } (i,j) \in P_2) \\ &\le \sum_{(i,j) \in P_1} \beta_j \cdot \mu(C_{ij} \cap A) && (\alpha_i \le \beta_j \text{ for any } (i,j) \in P_1) \\ &= \sum_{i=1}^n \sum_{j=1}^m \beta_i \cdot \mu(C_{ij} \cap A) && (\mu(C_{ij} \cap A) = 0 \text{ for any } (i,j) \in P_2) \\ &= \int_A g d\mu, \end{aligned}$$

which proves *ii*).

Suppose that f has two representations as the linear combination of a finite number of indicator functions

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i} = \sum_{i=1}^{m} \beta_i \cdot I_{B_i}.$$

Then, labeling $g = \sum_{i=1}^{m} \beta_i \cdot I_{B_i}$, $f = g$, so that, by the result proved above,

$$\begin{aligned} \int_A f d\mu &= \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A) \\ &= \int_A g d\mu = \sum_{j=1}^{m} \beta_j \cdot \mu(B_j \cap A) \end{aligned}$$

for any $A \in \mathcal{E}$. Therefore, the integral of f under μ over A is invariant to the representation of f, which tells us that the mapping $f \to \int_A f d\mu$ is a function for any $A \in \mathcal{E}$. This proves $i)$.

Finally, suppose once again that f, g have the canonical forms

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i} \quad \text{and} \quad g = \sum_{i=1}^{m} \beta_i \cdot I_{B_i},$$

nd let $c \in [0, +\infty)$.
Because $c \in [0, +\infty)$, $c\alpha_i \in [0, +\infty)$ for any $1 \leq i \leq n$, and

$$cf = \sum_{i=1}^{n} c\alpha_i \cdot I_{A_i}$$

is also a measurable simple function.
As above, define the disjoint sets $C_{ij} = A_i \cap B_j$ for any $1 \leq i \leq n$, $1 \leq j \leq m$. It was shown in lemma 2.11 that we can express

$$cf = \sum_{i=1}^{n} \sum_{j=1}^{n} c\alpha_i \cdot I_{C_{ij}} \quad \text{and} \quad g = \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j \cdot I_{C_{ij}},$$

so that

$$cf + g = \sum_{i=1}^{n} \sum_{j=1}^{m} (c\alpha_i + \beta_j) \cdot I_{C_{ij}}.$$

Furthermore, we showed in the proof of $ii)$ that

$$\int_A f d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \cdot \mu(C_{ij} \cap A)$$

and

$$\int_A g d\mu = \sum_{j=1}^{m} \beta_j \cdot \mu(B_j \cap A) = \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_i \cdot \mu(C_{ij} \cap A).$$

Therefore, for any $A \in \mathcal{E}$, by the definition of the integral,

$$\begin{aligned}
\int_A (cf+g) d\mu &= \sum_{i=1}^{n} \sum_{j=1}^{m} (c\alpha_i + \beta_j) \cdot \mu(C_{ij} \cap A) \\
&= c \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \cdot \mu(C_{ij} \cap A) + \sum_{i=1}^{n} \sum_{j=1}^{m} \beta_j \cdot \mu(C_{ij} \cap A) \\
&= c \cdot \int_A f d\mu + \int_A g d\mu.
\end{aligned}$$

Q.E.D.

We can now construct the integral of a non-negative measurable function with respect to μ over any measurable set A using the integral of simple functions. Let $\mathcal{E}^s$ be the set of all measurable simple functions.
Let $f \in \mathcal{E}_+$ be an arbitrary non-negative mesurable function. Then, for any $A \in \mathcal{E}$, we define the integral of f with respect to the measure μ over the measurable set A as

$$\int_A f d\mu = \sup_{g \leq f, g \in \mathcal{E}^s} \int_A g d\mu.$$

Note that this supremum is well defined on $[0, +\infty]$; if the set $\{\int_A g d\mu \mid g \leq f, g \in \mathcal{E}^s\}$ is bounded above, then by the least upper bound property of real numbers, it has a supremum in $[0, +\infty)$, while if it is unbounded, then its supremum is $+\infty$. Furthermore, this value is unique due to the uniqueness of the supremum, which means that the mapping $f \to \int_A f d\mu$ is a function.

Since simple measurable functions are also non-negative measurable functions, we seem to have two different definitions for the integral of a simple measurable function. However, we can easily show that the two definitions agree; letting f be a simple measurable function, denote the first definition by $(\int_A f d\mu)_1$ and the second by $(\int_A f d\mu)_2$.
For any $g \in \mathcal{E}^s$ such that $g \leq f$, by lemma 3.1 we have

$$\left(\int_A g d\mu\right)_1 \leq \left(\int_A f d\mu\right)_1,$$

so that $(\int_A f d\mu)_1$ is an upper bound of the set $\{(\int_A g d\mu)_1 \mid g \leq f, g \in \mathcal{E}_+\}$; this implies that

$$\left(\int_A f d\mu\right)_2 = \sup_{g \leq f, g \in \mathcal{E}^s} \left(\int_A g d\mu\right)_1 \leq \left(\int_A f d\mu\right)_1.$$

However, f is itself a simple measurable function that is less than or equal to f on E, so by definition of the supremum,

$$\left(\int_A f d\mu\right)_1 \leq \left(\int_A f d\mu\right)_2.$$

This shows us that $(\int_A f d\mu)_1 = (\int_A f d\mu)_2$, and that the two definitions agree for simple measurable functions.

The following are some useful properties of integrals of non-negative functions:

Theorem 3.2 Let $(E, \mathcal{E}, \mu)$ be a measure space, and $f, g \in \mathcal{E}_+$. The following hold true:

i) (Monotonicity) If $f \leq g$, then for any $A \in \mathcal{E}$, $\int_A f d\mu \leq \int_A g d\mu$.

ii) If $A, B \in \mathcal{E}$ and $A \subset B$, then $\int_A f d\mu \leq \int_B f d\mu$.

iii) For any $A \in \mathcal{E}$, if $f(x) = 0$ for any $x \in A$, then $\int_A f d\mu = 0$.

iv) For any $c \in [0, +\infty)$ and $A \in \mathcal{E}$, $c \cdot \int_A f d\mu = \int_A (cf) d\mu$.

v) If $\mu(A)=0$ for some $A\in\mathcal{E}$, then $\int_A f d\mu=0$.

vi) For any $A\in\mathcal{E}$, $\int_A f d\mu=\int_E (f\cdot I_A) d\mu$.

Proof) i) Choose any $A\in\mathcal{E}$. Suppose s is a simple measurable function such that $s\leq f$. Then, $s\leq g$ as well because $f\leq g$, meaning that

$$\int_A s d\mu \leq \int_A g d\mu.$$

This holds for any simple measurable function s such that $s\leq f$, so $\int_A g d\mu$ is an upper bound of the set $\{\int_A s d\mu \mid s\leq f, s\in\mathcal{E}^s\}$, and by the definition of the supremum, it follows that

$$\int_A f d\mu = \sup\{\int_A s d\mu \mid s\leq f, s\in\mathcal{E}^s\} \leq \int_A g d\mu.$$

ii) Choose $A,B\in\mathcal{E}$ such that $A\subset B$. For any simple measurable function $h=\sum_{i=1}^n \alpha_i\cdot I_{A_i}$, by definition

$$\int_A h d\mu = \sum_{i=1}^n \alpha_i\cdot\mu(A_i\cap A) \leq \sum_{i=1}^n \alpha_i\cdot\mu(A_i\cap B) = \int_B h d\mu,$$

where the inequality holds because $A_i\cap A\subset A_i\cap B$ for $1\leq i\leq n$ and measures are monotonic. Therefore, for any simple measurable function h such that $h\leq f$,

$$\int_A h d\mu \leq \int_B h d\mu \leq \int_B f d\mu,$$

so that $\int_B f d\mu$ is an upper bound of the set $\{\int_A s d\mu \mid s\leq f, s\in\mathcal{E}^s\}$ and thus

$$\int_A f d\mu = \sup\{\int_A s d\mu \mid s\leq f, s\in\mathcal{E}^s\} \leq \int_B f d\mu.$$

iii) Let $A\in\mathcal{E}$ and suppose $f(x)=0$ for any $x\in A$. Suppose s is a measurable simple function such that $s\leq f$. Then, $0\leq s(x)\leq f(x)=0$ and thus $s(x)=A$ for any $x\in A$.
Letting $s=\sum_{i=1}^n \alpha_i\cdot I_{A_i}$ for $\alpha_1,\cdots,\alpha_n\in(0,+\infty)$, this implies that $A_i\cap A=\emptyset$ for $1\leq i\leq n$; otherwise, if $A_i\cap A\neq\emptyset$ for some $1\leq i\leq n$, then letting $x\in A_i\cap A$, $s(x)=\alpha_i>0$, a contradiction. Therefore,

$$\int_A s d\mu = \sum_{i=1}^n \alpha_i\cdot\mu(A_i\cap A) = \sum_{i=1}^n \alpha_i\cdot\mu(\emptyset) = 0.$$

This holds for any simple measurable function s such that $s\leq f$, so $\{\int_A s d\mu \mid s\leq$

$f, s \in \mathcal{E}^s\} = \{0\}$ and as such

$$\int_A f d\mu = \sup\{\int_A s d\mu \mid s \leq f, s \in \mathcal{E}^s\} = 0.$$

iv) Choose any $c \in [0, +\infty)$ and $A \in \mathcal{E}$. If $c = 0$, then $(cf)(x) = cf(x) = 0$ for any $x \in E$, so that

$$\int_A (cf) d\mu = 0 = 0 \cdot \int_A f d\mu = c \cdot \int_A f d\mu$$

by result iii).

Suppose on the other hand that $c \in (0, +\infty)$. Then, for any simple measurable function s such that $s \leq f$, letting $s = \sum_{i=1}^n \alpha_i \cdot I_{A_i}$,

$$c \cdot s = \sum_{i=1}^n c\alpha_i \cdot I_{A_i},$$

where $c\alpha_i \in [0, +\infty)$ for any $1 \leq i \leq n$. It follows that $c \cdot s$ is a simple measurable function such that $c \cdot s \leq cf$, so

$$\begin{aligned} c \cdot \int_A s d\mu &= \int_A (cs) d\mu && \text{(lemma 3.1)} \\ &\leq \int_A (cf) d\mu. \end{aligned}$$

This holds for any simple measurable function s such that $s \leq f$, so

$$\begin{aligned} c \cdot \int_A f d\mu &= c \cdot \sup\{\int_A s d\mu \mid s \leq f, s \in \mathcal{E}^s\} \\ &= \sup\{c \cdot \int_A s d\mu \mid s \leq f, s \in \mathcal{E}^s\} && (c \text{ is positive}) \\ &\leq \int_A (cf) d\mu. \end{aligned}$$

Since $0 < c < +\infty$ and cf is also a non-negative measurable function, applying the above result implies that

$$\frac{1}{c} \cdot \int_A (cf) d\mu \leq \int_A \left(\frac{1}{c} cf\right) d\mu = \int_A f d\mu,$$

or $\int_A (cf) d\mu \leq c \cdot \int_A f d\mu$. Therefore,

$$\int_A (cf) d\mu = c \cdot \int_A f d\mu.$$

v) Suppose that $\mu(A) = 0$ for some $A \in \mathcal{E}$. Then, for any simple measurable function s such that $s \leq f$, letting $s = \sum_{i=1}^n \alpha_i \cdot I_{A_i}$, because $0 \leq \mu(A_i \cap A) \leq \mu(A) = 0$ implies

$\mu(A_i \cap A) = 0$ for $1 \le i \le n$,

$$\int_A s d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A) = 0.$$

This holds for any simple measurable function s such that $s \le f$, so $\{\int_A s d\mu \mid s \le f, s \in \mathcal{E}^s\} = \{0\}$ and as such

$$\int_A f d\mu = \sup\{\int_A s d\mu \mid s \le f, s \in \mathcal{E}^s\} = 0.$$

vi) For any $A \in \mathcal{E}$, first note that $g = f \cdot I_A$ is a non-negative measurable function because

$$g^{-1}([-\infty, a)) = \begin{cases} f^{-1}([-\infty, a)) \cap A & \text{if } a > 0 \\ \emptyset & \text{if } a \le 0 \end{cases} \in \mathcal{E}$$

for any $a \in \mathbb{Q}$.

Let s be a simple measurable function such that $s \le f$, and let $s = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$. Then,

$$s \cdot I_A = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i \cap A},$$

and as such

$$\int_E (s \cdot I_A) d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A) = \int_A s d\mu.$$

Since $s \cdot I_A \le f \cdot I_A$, we can see that

$$\int_A s d\mu = \int_E (s \cdot I_A) d\mu \le \int_E (f \cdot I_A)\, d\mu.$$

This holds for any simple measurable function s such that $s \le f$, so

$$\int_A f d\mu \le \int_E (f \cdot I_A)\, d\mu.$$

To show the reverse inequality, suppose s is a simple measurable function such that $s \le f \cdot I_A$. Then, for any $x \in A^c$, $s(x) = 0$. Letting $s = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$ for $\alpha_1, \cdots, \alpha_n \in (0, +\infty)$, this means that $A_i \subset A$ for $1 \le i \le n$; otherwise, if there exists an $1 \le i \le n$

such that $A_i \cap A^c \neq \emptyset$, then for any $x \in A_i \cap A^c$,

$$0 = s(x) = \alpha_i > 0,$$

a contradiction. Therefore,

$$\int_E s d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i) = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i \cap A) = \int_A s d\mu \leq \int_A f d\mu.$$

This holds for any simple measurable function s such that $s \leq f \cdot I_A$, so

$$\int_E (f \cdot I_A) d\mu \leq \int_A f d\mu,$$

from which we can conclude that

$$\int_E (f \cdot I_A) d\mu = \int_A f d\mu.$$

Q.E.D.

We can also see that two important properties hold:

Theorem 3.3 Let $(E, \mathcal{E}, \mu)$ be a measure space, and $f \in \mathcal{E}_+$. The following hold true:

i) (The Vanishing Property) For any $A \in \mathcal{E}$, $\int_A f d\mu = 0$ if and only if $\{x \in A \mid f(x) > 0\} = f^{-1}((0, +\infty]) \cap A$ has measure 0 under μ.

ii) (The Finiteness Property) For any $A \in \mathcal{E}$, if $\int_A f d\mu < +\infty$, then $\{x \in A \mid f(x) = +\infty\} = f^{-1}(\{+\infty\}) \cap A$ has measure 0 under μ.

Proof) i) Let $A \in \mathcal{E}$. Suppose $f^{-1}((0, +\infty]) \cap A$ has measure 0 under μ. In other words, letting $B = f^{-1}((0, +\infty]) \cap A = \{x \in A \mid f(x) > 0\}$, $\mu(B) = 0$. Note that

$$f \cdot I_A = f \cdot I_B + f \cdot I_{A \setminus B}.$$

By theorem 3.2,

$$\int_E (f \cdot I_B) d\mu = \int_B f d\mu = 0$$

because $\mu(B) = 0$, while

$$\left(f \cdot I_{A \setminus B}\right)(x) = f(x) \cdot I_{A \setminus B}(x) = 0$$

for any $x \in E$ because $f(x) = 0$ if $x \in A \setminus B$, which implies that

$$f \cdot I_A = f \cdot I_B.$$

Therefore,

$$\int_A f d\mu = \int_E (f \cdot I_A) d\mu = \int_E (f \cdot I_B) d\mu = 0.$$

Conversely, suppose that $\int_A f d\mu = 0$. For any $n \in N_+$, define $B_n = f^{-1}((0, +\infty]) \cap A = \{x \in A \mid f(x) > \frac{1}{n}\}$, and let $s_n = \frac{1}{n} \cdot I_{B_n}$. Then, $0 \leq s \leq f$ for any $x \in E$, and

$$\int_A s d\mu = \frac{1}{n}\mu(A \cap B_n) = \frac{1}{n}\mu(B_n).$$

By the definition of the integral of non-negative functions,

$$0 \leq \frac{1}{n}\mu(B_n) \leq \int_A f d\mu = 0,$$

so that $\frac{1}{n}\mu(B_n) = 0$, or $\mu(B_n) = 0$.
This holds for any $n \in N_+$, and note that $\bigcup_n B_n = B$. Therefore, by sequential continuity,

$$\mu(B) = \lim_{n \to \infty} \mu(B_n) = 0.$$

ii) Let $A \in \mathcal{E}$, and suppose that $\int_A f d\mu < +\infty$. To show that *ii*) holds, assume the contrary and suppose $\mu(B) > 0$ for $B = f^{-1}(\{+\infty\}) \cap A = \{x \in A \mid f(x) = +\infty\}$. Define the sequence of simple measurable functions $\{s_n\}_{n \in N_+}$ as $s_n = n \cdot I_B$. Then, $0 \leq s_n \leq f$ for any $n \in N_+$, and by the definition of the integral of a non-negative function,

$$\int_A s_n d\mu = n \cdot \mu(A \cap B) = n \cdot \mu(B) \leq \int_A f d\mu.$$

If $\mu(B) = +\infty$, we have the contradiction $+\infty \leq \int_A f d\mu < +\infty$. If $\mu(B) \in (0, +\infty)$, this contradicts the Archimedean property of the real numbers, since $\int_A f d\mu \in (0, +\infty)$ as well. Therefore, $\mu(B) = 0$.

Q.E.D.

3.2 The Monotone Convergence Theorem

This section is dedicated to one of the most fundamental results of measure theory, which allows us to exchange limits and integrals for increasing sequences of functions.
We first need the following preliminary result:

Lemma 3.4 Let $(E, \mathcal{E}, \mu)$ be a measure space, and f a simple $\mathcal{E}$-measurable function. Then, for any sequence of disjoint measurable sets $\{A_n\}_{n \in N_+} \subset \mathcal{E}$, letting $A = \bigcup_n A_n$,

$$\int_A f d\mu = \sum_{n=1}^{\infty} \int_{A_n} f d\mu,$$

where the limit on the right exists in $[0, +\infty]$.

Proof) Let $f = \sum_{i=1}^{n} \alpha_i \cdot I_{B_i}$ be the canonical form of f. Then, note that

$$\begin{aligned}
\int_A f d\mu &= \sum_{i=1}^{n} \alpha_i \cdot \mu(B_i \cap A) \\
&= \sum_{i=1}^{n} \alpha_i \cdot \mu\left(\bigcup_n (A_n \cap B_i)\right) \\
&= \sum_{i=1}^{n} \alpha_i \cdot \left[\sum_{n=1}^{\infty} \mu(A_n \cap B_i)\right] && \text{(Countable Additivity of } \mu) \\
&= \sum_{n=1}^{\infty} \left(\sum_{i=1}^{n} \alpha_i \cdot \mu(A_n \cap B_i)\right) \\
&= \sum_{n=1}^{\infty} \int_{A_n} f d\mu.
\end{aligned}$$

Q.E.D.

We are now ready to prove the main theorem of this section.

Theorem 3.5 (The Monotone Convergence Theorem, MCT)
Let $(E,\mathcal{E},\mu)$ be a measure space, and $\{f_n\}_{n\in N_+}$ a sequence of increasing non-negative $\mathcal{E}$-measurable functions. Then, the pointwise limit $f\in\mathcal{E}_+$ of $\{f_n\}_{n\in N_+}$ exists and

$$\lim_{n\to\infty}\int_E f_n d\mu = \int_E f d\mu,$$

where the limit on the left hand side exists in $[0,+\infty]$.

Proof) For any $x\in E$, since $\{f_n(x)\}_{n\in N_+}$ is an increasing sequence in $[0,+\infty]$, if $\{f_n(x)\}_{n\in N_+}$ is bounded then $f_n(x)\nearrow f_x$ for some $f_x\in[0,+\infty)$, while if $\{f_n(x)\}_{n\in N_+}$ is unbounded, then $f_n(x)\nearrow f_x=+\infty$. Defining $f:E\to[-\infty,+\infty]$ as $f(x)=f_x$ for any $x\in E$, $f_n\nearrow f$ pointwise, and because measurability is preserved over limits, f is non-negative and $\mathcal{E}$-measurable.

For any $n\in N_+$, $f_n\le f_{n+1}\le f$. Thus, by the monotonicity of integration, $\int_E f_n d\mu\le\int_E f_{n+1}d\mu\le\int_E f d\mu$. This holds for any $n\in N_+$, so $\{\int_E f_n d\mu\}_{n\in N_+}$ is an increasing sequence in $[0,+\infty]$ that is bounded above by $\int_E f d\mu$; it follows that

$$\lim_{n\to\infty}\int_E f_n d\mu\le\int_E f d\mu,$$

where the limit on the left hand side exists in $[0,+\infty]$ because $\{\int_E f_n d\mu\}_{n\in N_+}$ is increasing.

It remains to be seen that the reverse inequality holds, or that $\alpha=\lim_{n\to\infty}\int_E f_n d\mu\ge\int_E f d\mu$ holds. To this end, put $c\in(0,1)$ and let s be a simple measurable function such that $0\le s\le f$. Define

$$A_n=\{x\in E\mid f_n(x)\ge c\cdot s(x)\}.$$

By theorem 2.7, f_n and $c\cdot s$ are both non-negative measurable functions and, being a set of the form introduced in theorem 2.7 *v*), A_n is a measurable set. In addition, $A_n\subset A_{n+1}$ for any $n\in N_+$ because if $x\in A_n$, then $c\cdot s(x)\le f_n(x)\le f_{n+1}(x)$ and $x\in A_{n+1}$.
For any $x\in E$, if $s(x)>0$, then $0\le c\cdot s(x)<s(x)\le f(x)$, and because $\lim_{n\to\infty}f_n(x)=f(x)$, there exists an $N\in N_+$ such that

$$|f(x)-f_n(x)|<f(x)-c\cdot s(x),$$

and in particular

$$c\cdot s(x)<f_n(x),$$

for any $n\ge N$. In other words, $x\in A_n$ for any $n\ge N$.
On the other hand, if $s(x)=0$, then $x\in A_n$ for any $n\in N_+$ trivially.

Therefore, defining $A = \bigcup_n A_n$, we can see that $E \subset A$, or $E = A = \bigcup_n A_n$.

For any $n \in N_+$, by monotonicity, and the fact that $f_n \geq f_n \cdot I_{A_n}$, we can see that

$$\int_E f_n d\mu \geq \int_E (f_n \cdot I_{A_n}) d\mu = \int_{A_n} f_n d\mu \geq \int_{A_n} (c \cdot s) d\mu = c \cdot \int_{A_n} s d\mu$$

by the results in theorem 3.2. By lemma 3.4 and sequential continuity, it follows that

$$\alpha = \lim_{n \to \infty} \int_E f_n d\mu \geq c \cdot \int_A s d\mu.$$

This holds for any simple measurable function $s \leq f$, so

$$c \cdot \int_E f d\mu = c \cdot \sup\{\int_E g d\mu \mid g \in \mathcal{E}^s, g \leq f\} \leq \alpha.$$

This in turn holds for any $c \in (0,1)$, so

$$\frac{n-1}{n} \cdot \int_E f d\mu \leq \alpha$$

for any $n \geq 2$; taking $n \to \infty$ on both sides, we can now see that

$$\int_E f d\mu \leq \alpha.$$

This completes the proof.
Q.E.D.

3.2.1 The Linearity of Integration

There are numerous applications of the MCT: first, the linearity of the integration of non-negative functions:

First Corollary to the MCT (The Linearity of Integration)
Let $(E, \mathcal{E}, \mu)$ be a measure space, and f, g non-negative measurable functions. Then,

$$\int_E (f+g)d\mu = \int_E f d\mu + \int_E g d\mu.$$

Proof) Let $\{s_n\}_{n \in N_+}$ and $\{h_n\}_{n \in N_+}$ be increaing sequences of simple measurable functions increasing to f and g respectively. The sum of simple functions is also a simple function, and measurability is preserved over the sum of non-negative functions, so $\{s_n + h_n\}_{n \in N_+}$ is a sequence of simple measurable functions increasing to the non-negative measurable function $f+g$.
Because the integration of simple functions is linear,

$$\int_E (s_n + h_n)d\mu = \int_E s_n d\mu + \int_E h_n d\mu$$

for any $n \in N_+$. Note that $\{s_n + h_n\}_{n \in N_+}$, $\{s_n\}_{n \in N_+}$ and $\{h_n\}_{n \in N_+}$ are increasing sequences of non-negative measurable functions increasing to f, g, and $f+g$ respectively. By the MCT, we can see that

$$\begin{aligned}
\int_E (f+g)d\mu &= \lim_{n\to\infty} \int_E (s_n + h_n)d\mu \\
&= \lim_{n\to\infty} \int_E s_n d\mu + \lim_{n\to\infty} \int_E h_n d\mu \\
&= \int_E f d\mu + \int_E g d\mu.
\end{aligned}$$

Q.E.D.

3.2.2 The MCT for Series

The next corollary concerns the series version of the MCT:

Second Corollary to the MCT (The MCT for Series)
Let $(E, \mathcal{E}, \mu)$ be a measure space, and $\{f_n\}_{n\in N_+}$ a sequence of non-negative measurable functions. Then, $f = \sum_{n=1}^{\infty} f_n$ is a well-defined non-negative measurable function, and

$$\sum_{n=1}^{\infty} \int_E f_n d\mu = \int_E f d\mu,$$

where the series on the left hand side exists in $[0, +\infty]$.

Proof) For any $n \in N_+$, define

$$g_n = \sum_{i=1}^{n} f_i,$$

which is a well defined non-negative numerical function because $f_i(x)$ is non-negative for any $x \in E$. The finite sum of non-negative measurable functions preserves measurability, so g_n is measurable. $\{g_n\}_{n\in N_+}$ is then an increasing sequence of $\mathcal{E}$-measurable functions, so that the pointwise limit f of $\{g_n\}_{n\in N_+}$ exists and is a non-negative measurable function. By definition, for any $x \in E$

$$f(x) = \lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \sum_{i=1}^{n} f_i(x) = \sum_{n=1}^{\infty} f_n(x).$$

Therefore, $f = \sum_{n=1}^{\infty} f_n$ is a well-defined non-negative measurable function.

Now note that, for any $n \in N_+$,

$$\int_E g_n d\mu = \sum_{i=1}^{n} \int_E f_i d\mu$$

by the linearity of the integration of non-negative functions proved above. Each $\int_E f_n d\mu \in [0, +\infty]$, so it follows that the sequence $\{\sum_{i=1}^{n} \int_E f_i d\mu\}_{n\in N_+}$ is increasing in $[0, +\infty]$; thus, it increases to the limit

$$\lim_{n\to\infty} \sum_{i=1}^{n} \int_E f_i d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu \in [0, +\infty].$$

$\{g_n\}_{n\in N_+}$ is a sequence of non-negative functions increasing to f, so by the MCT,

$$\int_E f d\mu = \lim_{n\to\infty} \int_E g_n d\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int_E f_i d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Q.E.D.

3.2.3 Series as Integrals

Let c the counting measure on $(N_+, 2^{N_+})$. Then, for any non-negative function f on N_+, f is measurable because N_+ is countable, and the sequence $\{s_n\}_{n\in N_+}$ of simple functions defined as

$$s_n = \sum_{i=1}^{n} f(i) I_{\{i\}}$$

increases to f. It follows that the integral of f is

$$\int_{N_+} f dc = \lim_{n\to\infty} \sum_{i=1}^{n} f(i)\cdot c(\{i\}) = \lim_{n\to\infty} \sum_{i=1}^{n} f(i) = \sum_{n=1}^{\infty} f(n) \in [0, +\infty].$$

This observation motivates the treatment of series as integrals, and leads to the following corollary of the MCT:

Third Corollary to the MCT (Series as Integrals)
Let $\{a_{mn}\}_{n,m\in N_+} \in [0, +\infty]$ be a sequence of non-negative numbers in the extended real number system. The following results hold:

i) If $a_{nm} \le a_{n+1,m}$ for any $n, m \in N_+$, then there exist $b_m = \lim_{n\to\infty} a_{nm}$ in $[0, +\infty]$ for any $m \in N_+$, and

$$\lim_{n\to\infty} \sum_{m=1}^{\infty} a_{nm} = \sum_{m=1}^{\infty} b_m.$$

ii) The order of summation can be interchanged:

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{nm} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{nm}.$$

Proof) i) For any $n \in N_+$, define the function $f_n : N_+ \to [0, +\infty]$ as

$$f_n(m) = a_{nm}$$

for any $m \in N_+$. Then, letting c be the counting measure on $(N_+, 2^{N_+})$,

$$\int_{N_+} f_n dc = \sum_{m=1}^{\infty} f_n(m) = \sum_{m=1}^{\infty} a_{nm},$$

and $f_n \le f_{n+1}$ for any $n \in N_+$ because

$$f_n(m) = a_{nm} \le a_{n+1,m} = f_{n+1}(m)$$

for any $m \in N_+$ by assumption. Thus, by the MCT, we know that $f = \lim_{n\to\infty} f_n$ exists, that is, for any $m \in N_+$ there exists a $b_m = f(m) = \lim_{n\to\infty} f_n(m) =$

$\lim_{n\to\infty} a_{nm}$, and

$$\lim_{n\to\infty}\sum_{m=1}^{\infty} a_{nm} = \lim_{n\to\infty}\int_{N_+} f_n dc = \int_{N_+} f dc = \sum_{m=1}^{\infty} b_m.$$

ii) Letting c be the counting measure on N_+, define $f_n : N_+ \to [0,+\infty]$ as

$$f_n(m) = a_{nm}$$

for any $m \in N_+$. Then, as above,

$$\int_{N_+} f_n dc = \sum_{m=1}^{\infty} f_n(m) = \sum_{m=1}^{\infty} a_{nm},$$

for any $n \in N_+$.

$\{f_n\}_{n\in N_+}$ is a sequence of non-negative functions on N_+, so that, by the MCT for series, the function $f : N_+ \to [0,+\infty]$ defined as

$$f(m) = \sum_{n=1}^{\infty} f_n(m) = \sum_{n=1}^{\infty} a_{nm}$$

for any $m \in N_+$ is well-defined in $[0,+\infty]$, and

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{nm} = \sum_{n=1}^{\infty}\int_{N_+} f_n dc = \int_{N_+} f dc = \sum_{m=1}^{\infty} f(m) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{nm}.$$

Q.E.D.

3.2.4 Indefinite Integrals and Radon-Nikodym Derivatives

We can also formulate an extension of lemma 3.4 to non-negative functions, not just simple functions.

Fourth Corollary to Theorem 3.5 (The Indefinite Integral of a Function) Let $(E, \mathcal{E}, \mu)$ be a measure space, and f a non-negative measurable function. Then, the function $v : \mathcal{E} \to [0, +\infty]$ defined as

$$v(A) = \int_A f d\mu$$

is a measure on $(E, \mathcal{E})$. Moreover, for any $g \in \mathcal{E}_+$,

$$\int_E g dv = \int_E g f d\mu.$$

Proof) Defining $\{s_n\}_{n \in N_+}$ as the sequence of simple measurable functions increasing to the non-negative measurable function f, note that, for any $A \in \mathcal{E}$, $s_n \cdot I_A$ is a simple measurable function for any $n \in N_+$ (the product of simple functions is simple, and measurability of real functions are preserved under products), and that $s_n \cdot I_A \nearrow f \cdot I_A$. We will verify that v satisfies the conditions for a measure:

i) $v(\emptyset) = \int_\emptyset f d\mu = 0$ by theorem 3.2.

ii) For any sequence of disjoint measurable sets $\{A_m\}_{m \in N_+} \subset \mathcal{E}$, let $A = \bigcup_m A_m$.

$$\begin{aligned} v(A) = \int_A f d\mu = \int_E (f \cdot I_A) d\mu & \qquad \text{(Theorem 3.2)} \\ = \lim_{n \to \infty} \int_E (s_n \cdot I_A) d\mu & \qquad \text{(The MCT)} \\ = \lim_{n \to \infty} \int_A s_n d\mu & \qquad \text{(Theorem 3.2)} \\ = \lim_{n \to \infty} \sum_{m=1}^{\infty} \int_{A_m} s_n d\mu. & \qquad \text{(Lemma 3.4)} \end{aligned}$$

Define the double sequence $\{a_{nm}\}_{n,m \in N_+}$ in $[0, +\infty]$ as

$$a_{nm} = \int_{A_m} s_n d\mu$$

for any $n, m \in N_+$. Then, because $a_{nm} \leq a_{n+1,m}$ for any $n, m \in N_+$ by the monotonicity of integration, by the third corollary to the MCT we have

$$\begin{aligned}\lim_{n\to\infty}\sum_{m=1}^{\infty} a_{nm} &= \sum_{m=1}^{\infty}\left(\lim_{n\to\infty} a_{nm}\right)\\ &= \sum_{m=1}^{\infty}\left(\lim_{n\to\infty}\int_{A_m} s_n d\mu\right)\\ &= \sum_{m=1}^{\infty}\int_{A_m} f d\mu && \text{(The MCT)}\\ &= \sum_{m=1}^{\infty} v(A_m).\end{aligned}$$

By definition, v is a measure on $(E,\mathcal{E})$.

Let $g \in \mathcal{E}_+$ be a simple measurable function with canonical form

$$g = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}.$$

Then,

$$\begin{aligned}\int_E g dv = \sum_{i=1}^{n}\alpha_i\cdot v(A_i) = \sum_{i=1}^{n}\alpha_i\cdot\int_E (f\cdot I_{A_i})\, d\mu\\ &= \int_E f\cdot\left(\sum_{i=1}^{n}\alpha_i\cdot I_{A_i}\right) d\mu && \text{(The Linearity of Integration)}\\ &= \int_E fg d\mu.\end{aligned}$$

Now let $g \in \mathcal{E}_+$ be an arbitrary non-negative measurable function. Letting $\{s_n\}_{n\in N_+}$ be a sequence of simple measurable functions increasing to g, $\{fs_n\}_{n\in N_+}$ is a sequence of non-negative measurable functions (the product of non-negative measurable functions are measurable) increasing to fg. Thus, by repeated applications of the MCT,

$$\begin{aligned}\int_E g dv &= \lim_{n\to\infty}\int_E s_n dv && \text{(The MCT)}\\ &= \lim_{n\to\infty}\int_E (f\cdot s_n)\, d\mu && \text{(The preceding result)}\\ &= \int_E g f d\mu. && \text{(The MCT)}\end{aligned}$$

Q.E.D.

The measure v defined above is called the indefinite integral of f under μ. When μ is the Lebesgue measure on $\mathbb{R}$, note that v satisfies

$$v((-\infty,x]) = \int_{-\infty}^{x} f(y)dy$$

for any $x \in \mathbb{R}$, so that

$$\int_{a}^{b} f(y)dy = v((-\infty,b]) - v((-\infty,a])$$

for any $-\infty < a < b < +\infty$. In this case, v is a function that allows us to compute the value of the definite integral $\int_a^b f(y)dy$ for any open interval (a,b); it is in this sense that v is called the "indefinite" integral.

Alternatively, if, for two measures μ, v on $(E,\mathcal{E})$, there exists a non-negative measurable function $f \in \mathcal{E}_+$ such that

$$v(A) = \int_A f d\mu$$

for any $A \in \mathcal{E}$, then f is called the Radon-Nikodym derivative, or the density, of v with respect to μ.
In probability theory, any probability mass function or probability density function of a univariate random variable is the Radon-Nikodym derivative of the associated distribution with respect to the counting measure on $\mathbb{N}$ or the Lebesgue measure on $\mathbb{R}$. This will be made clearer when the distribution of a random variable is defined.

For two measures μ, v on $(E,\mathcal{E})$, we say that v is absolutely continuous with respect to μ, denoted $v << \mu$, if, for any $A \in \mathcal{E}$ such that $\mu(A) = 0$, $v(A) = 0$ holds as well.
Note that, if there exists a Radon-Nikodym derivative f of v with respect to μ, then $v << \mu$, since if $\mu(A) = 0$ for some $A \in \mathcal{E}$, then

$$v(A) = \int_A f d\mu = 0$$

by theorem 3.2.
There also exists a converse to this statement: if v, μ are σ-finite and $v << \mu$, then there exists a Radon-Nikodym derivative f of v with respect to μ. This is called the Radon-Nikodym theorem, and it is used to derive many fundamental results in probability theory, including the existence of conditional expectations and densities of discrete and continuous random variables.
The Radon-Nikodym theorem and more are studied in detail in section 5 of chapter 6 using Hilbert space techniques.

3.2.5 Fatou's Lemma

This corollary to the MCT will be used to derive many useful results later on for the integral of complex measurable functions. It is also of independent interest and is used as the linchpin in many proofs.

Fifth Corollary to the MCT (Fatou's Lemma)
Let $(E, \mathcal{E}, \mu)$ be a measure space, and $\{f_n\}_{n \in N_+}$ a sequence of non-negative measurable functions. Then, letting $f = \liminf_{n\to\infty} f_n$,

$$\int_E f d\mu \leq \liminf_{n\to\infty} \int_E f_n d\mu.$$

Proof) For any $n \in N_+$, define

$$g_n = \inf_{k \geq n} f_k;$$

g_n is non-negative because f_k are all bounded below by 0, and because measurability is preserved under infimums, g_n is also measurable. Furthermore, $g_n \leq g_{n+1}$. By the definition of the limit inferior,

$$f = \liminf_{n\to\infty} f_n = \sup_{n \in N_+} \left(\inf_{k \geq n} f_k \right) = \lim_{n\to\infty} g_n.$$

Thus, $\{g_n\}_{n \in N_+}$ is a sequence of non-negative measurable functions that increases to f; by the MCT,

$$\int_E f d\mu = \lim_{n\to\infty} \int_E g_n d\mu.$$

For any $n \in N_+$ and $k \geq n$, because $g_n = \inf_{m \geq n} f_m \leq f_k$, by the monotonicity of integration

$$\int_E g_n d\mu \leq \int_E f_k d\mu,$$

and since this holds for any $k \geq n$,

$$\int_E g_n d\mu \leq \inf_{k \geq n} \int_E f_k d\mu.$$

Therefore,

$$\int_E f d\mu = \lim_{n\to\infty} \int_E g_n d\mu \leq \lim_{n\to\infty} \left(\inf_{k \geq n} \int_E f_k d\mu \right) = \liminf_{n\to\infty} \int_E f_n d\mu.$$

Q.E.D.

3.3 Integration of Numerical and Complex Functions

3.3.1 Integration of Numerical Functions

Let $(E, \mathcal{E}, \mu)$ be a measure space, and $f : E \to [-\infty, +\infty]$ a numerical function. We showed earlier that f is $\mathcal{E}$-measurable if and only if its positive and negative parts f^+ and f^- are non-negative $\mathcal{E}$-measurable functions. The absolute value of f, $|f|$, is defined as

$$|f| = f^+ + f^-.$$

Suppose $f \in \mathcal{E}$. Then, since $|f|$ is the sum of two non-negative $\mathcal{E}$-measurable functions, it is also non-negative and $\mathcal{E}$-measurable.

Suppose there exists a $\mathcal{E}$-measurable numerical function f such that

$$\int_E f^{\pm} d\mu < +\infty.$$

Then, we say that f is μ-integrable, and define the integral of f over any $A \in \mathcal{E}$ as

$$\int_A f d\mu = \int_A f^+ d\mu - \int_A f^- d\mu.$$

The value above is well-defined in $\mathbb{R}$ because they are both finite values (their finiteness follows from the fact that the integral of a non-negative function over A is smaller than or equal to its integral over E by the monotonicity of integration).

Sometimes, we extend the definition of an integral of a numerical function to include $-\infty$ and $+\infty$. Specifically, given a measurable numerical function f, if either $\int_E f^+ d\mu < +\infty$ or $\int_E f^- d\mu < +\infty$, we define the integral of f over any $A \in \mathcal{E}$ as

$$\int_A f d\mu = \int_A f^+ d\mu - \int_A f^- d\mu \in [-\infty, +\infty],$$

since the difference here is well-defined due to the finiteness of one of the terms. In this case, we say that the integral of f over E exists in the ”extended” sense.

Consider a $\mathcal{E}$-measurable numerical function f. Since f is measurable, so is $|f|$; being a non-negative measurable function, the integral $\int_E |f| d\mu$ of $|f|$ is well-defined.
Suppose $\int_E |f| d\mu < +\infty$. Then, by the lineatiy of integration, because $|f| = f^+ + f^-$,

$$\int_E f^+ d\mu + \int_E f^- d\mu = \int_E |f| d\mu < +\infty.$$

$\int_E f^+ d\mu$ and $\int_E f^- d\mu$, being integrals of non-negative functions, take values in $[0, +\infty]$, so the fact that their sums are finite means that $\int_E f^+ d\mu$ and $\int_E f^- d\mu$ themselves are also finite. Therefore, f is μ-integrable by definition.
Conversely, suppose that f is μ-integrable. Then,$\int_E f^+ d\mu$ and $\int_E f^- d\mu$ take values in $[0, +\infty)$ by definition, and by the linearity of integration,

$$\int_E |f| d\mu = \int_E f^+ d\mu + \int_E f^- d\mu < +\infty.$$

We have seen above that f is μ-integrable if and only if the μ-integral of $|f|$ over E is integrable; since $|f|$ is its own positive part and its negative part is 0, this is equivalent to saying that $|f|$ is μ-integrable. Therefore, f is μ-integrable if and only if $|f|$ is μ-integrable.

The following are some elementary properties of the integrals of integrable functions:

Theorem 3.6 Let $(E, \mathcal{E}, \mu)$ be a measure space, and $f, g : E \to [-\infty, +\infty]$ numerical functions that are $\mathcal{E}$-measurable. Then, the following hold true:

i) If f is μ-integrable, then for any $A \in \mathcal{E}$, $f \cdot I_A$ is also μ-integrable and

$$\int_A f d\mu = \int_E (f \cdot I_A)\, d\mu.$$

ii) (Monotonicity) If f, g are μ-integrable and $f \leq g$, then $\int_E f d\mu \leq \int_E g d\mu$.

iii) (Linearity) For any $a \in \mathbb{R}$, if f, g are μ-integrable and real-valued, then $af + g$ is a $\mathcal{E}$-measurable and μ-integrable real-valued function such that

$$\int_E (af + g) d\mu = a \cdot \int_E f d\mu + \int_E g d\mu.$$

iv) If f is μ-integrable, then

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Proof) i) Let f be an $\mathcal{E}$-measurable μ-integrable function and $A \in \mathcal{E}$. Then, because $|f I_A| \leq |f|$,

$$\int_E |f I_A| d\mu \leq \int_E |f| d\mu < +\infty$$

by the monotonicity of integration, indicating that fI_A is also μ-integrable. Note that, because $\int_A g d\mu = \int_E (g \cdot I_A)\, d\mu$ for any non-negative measurable function g, and $(fI_A)^{\pm} = f^{\pm} I_A$, we have

$$\begin{aligned}
\int_A f d\mu &= \int_A f^+ d\mu - \int_A f^- d\mu && \text{(Definition)} \\
&= \int_E \left(f^+ I_A\right) d\mu - \int_E (f^- I_A)\, d\mu && \text{(theorem 3.2)} \\
&= \int_E (fI_A)^+ d\mu - \int_E (fI_A)^- d\mu \\
&= \int_E (fI_A)\, d\mu, && \text{(Definition of the Integral)}
\end{aligned}$$

where the rightmost integral exists becasue fI_A is integrable. Therefore, like with non-negative functions,

$$\int_A f d\mu = \int_E (f \cdot I_A)\, d\mu.$$

We will use this characterization from now on and focus on integrals over the entire set.

ii) Suppose $f \leq g$. Then, $f^+ - f^- \leq g^+ - g^-$, and adding $f^- + g^-$ to both sides yields the inequality

$$f^+ + g^- \leq g^+ + f^-.$$

Both $f^+ + g^-$ and $g^+ + f^-$ are non-negative measurable functions, so by the monotonicity and linearity of integration of non-negative functions,

$$\begin{aligned}
\int_E f^+ d\mu + \int_E g^- d\mu &= \int_E (f^+ + g^-) d\mu \\
&\leq \int_E (g^+ + f^-) d\mu = \int_E g^+ d\mu + \int_E f^- d\mu.
\end{aligned}$$

Since each of the integrals $\int_E f^+ d\mu$, $\int_E g^- d\mu$, $\int_E g^+ d\mu$, $\int_E f^- d\mu$ above are finite by the definition of μ-integrability, we can rearrange the terms so that

$$\int_E f^+ d\mu - \int_E f^- d\mu \leq \int_E g^+ d\mu - \int_E g^- d\mu.$$

By the definition of the integral of integrable functions, we finally have

$$\int_E f d\mu \leq \int_E g d\mu.$$

iii) af is $\mathcal{E}$-measurable, and because af and g are both real-valued, $af + g$ is $\mathcal{E}$-measurable as well. We proceed step by step.

First, assume that $a \in [0, +\infty)$. Then,

$$\begin{aligned}(af)^+ &= \max(af, 0) = a \cdot \max(f, 0) = af^+ \quad \text{and} \\ (af)^- &= -\min(af, 0) = a \cdot (-\min(f, 0)) = af^-,\end{aligned}$$

so that

$$\int_E (af)^\pm d\mu = a \cdot \int_E f^\pm d\mu < +\infty$$

by the linearity of integration of non-negative functions. By definition, af is μ-integrable, and

$$\int_E af d\mu = \int_E (af)^+ d\mu - \int_E (af)^- d\mu = a \cdot \left(\int_E f^+ d\mu - \int_E f^- d\mu \right) = a \cdot \int_E f d\mu.$$

On the other hand, suppose $a < 0$. Then,

$$\begin{aligned}(af)^+ &= \max(af, 0) = a \cdot \min(f, 0) = (-a)f^- = |a| f^- \quad \text{and} \\ (af)^- &= -\min(af, 0) = (-a) \cdot \max(f, 0) = (-a)f^+ = |a| f^+,\end{aligned}$$

so that

$$\int_E (af)^\pm d\mu = |a| \cdot \int_E f^\mp d\mu < +\infty$$

by the linearity of integration of non-negative functions. By definition, af is μ-integrable, and

$$\int_E af d\mu = \int_E (af)^+ d\mu - \int_E (af)^- d\mu = (-a) \cdot \left(\int_E f^- d\mu - \int_E f^+ d\mu \right) = a \cdot \int_E f d\mu.$$

In any case, af is μ-integrable and $\int_E af d\mu = a \cdot \int_E f d\mu$.

Now denote $h = af + g$. Since

$$(af + g)^+ \leq (af)^+ + g^+ \quad \text{and} \quad (af + g)^- \leq (af)^- + g^-,$$

by the monotonicity and linearity of integration of positive functions,

$$\int_E h^+ d\mu \leq \cdot \int_E (af)^+ d\mu + \int_E g^+ d\mu < +\infty, \int_E h^- d\mu \quad \leq \cdot \int_E (af)^- d\mu + \int_E g^- d\mu < +\infty,$$

where the inequalities follow from the μ-integrability of af and g. By definition, h is μ-integrable. To obtain the μ-integral of h over E, note that

$$(af)^+ - (af)^- + g^+ - g^- = h = h^+ - h^-.$$

Rearranging terms yields

$$(af)^+ + g^+ + h^- = h^+ + (af)^- + g^-.$$

The functions on both sides are non-negative and measurable, so by the linearity of integration of non-negative functions,

$$\int_E (af)^+ d\mu + \int_E g^+ d\mu + \int_E h^- d\mu = \int_E (af)^- d\mu + \int_E g^- d\mu + \int_E h^+ d\mu.$$

All the terms involved are finite by integrability, so rearranging terms once again yields

$$\begin{aligned}\int_E af d\mu + \int_E g d\mu &= \int_E (af)^+ d\mu - \int_E (af)^- d\mu + \int_E g^+ d\mu - \int_E g^- d\mu \\ &= \int_E h^+ d\mu - \int_E h^- d\mu = \int_E h d\mu.\end{aligned}$$

We saw above that $\int_E af d\mu = a \cdot \int_E f d\mu$, so

$$a \cdot \int_E f d\mu + \int_E g d\mu = \int_E af d\mu + \int_E g d\mu = \int_E h d\mu.$$

iv) This follows very easily; since $\int_E f^- d\mu, \int_E f^+ d\mu \in [0, +\infty)$,

$$\left| \int_E f d\mu \right| = \left| \int_E f^+ d\mu - \int_E f^- d\mu \right| \le \int_E f^+ d\mu + \int_E f^- d\mu = \int_E |f| d\mu,$$

where the last equality follows by the linearity of integration.

Q.E.D.

3.3.2 Integration of Complex Functions

A Characterization of Measurability for Complex Functions

Let $(E,\mathcal{E})$ be a measurable space, and $f : E \to \mathbb{C}$ a function defined on E that maps into the complex plane $\mathbb{C}$. Let $\tau_{\mathbb{C}}$ be the standard (euclidean) topology on $\mathbb{C}$, and define $\mathcal{B}(\mathbb{C})$ as the Borel σ-algebra generated by $\tau_{\mathbb{C}}$.
f is then measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{C})$ if $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{B}(\mathbb{C})$. We can obtain a simpler characterization of measurability if we rely on the fact that $\mathbb{C}$ is homeomorphic to $\mathbb{R}^2$.

Formally, define $\phi : \mathbb{C} \to \mathbb{R}^2$ as

$$\phi(z) = (Re(z), Im(z))$$

for any $z \in \mathbb{C}$. We saw in chapter 1 that ϕ is a homeomorphism between $(\mathbb{C}, \tau_{\mathbb{C}})$ and $(\mathbb{R}^2, \tau_{\mathbb{R}}^2)$. To facilitate our subsequent proofs, recall that $\tau_{\mathbb{R}}$ is generated by some countable base $\mathbb{B}^1$, and as such that $\tau_{\mathbb{R}}^2$ is generated by the countable base

$$\mathbb{B}^2 = \{A \times B \mid A, B \in \mathbb{B}^1\}.$$

By lemma 2.2, $\mathbb{B}^1$ generates $\mathcal{B}(\mathbb{R})$ and $\mathbb{B}^2$ generates the Borel σ-algebra on $\mathbb{R}^2$.

Let $f : E \to \mathbb{C}$ be a complex function, and define $h = \phi \circ f = (Re(f), Im(f))$.
Suppose that $Re(f)$ and $Im(f)$ are $\mathcal{E}$-measurable real functions. Then,

$$h^{-1}(A \times B) = (Re(f))^{-1}(A) \cap (Im(f))^{-1}(B) \in \mathcal{E}$$

for any $A \times B \in \mathbb{B}^2$. Since $\mathbb{B}^2$ generates the Borel σ-algebra on $\mathbb{R}^2$, it follows that h is measurable relative to $\mathcal{E}$ and the Borel σ-algebra on $\mathbb{R}^2$. Because $f = \phi^{-1} \circ h$, and the continuity of $\phi^{-1} : \mathbb{R}^2 \to \mathbb{C}$ relative to $\tau_{\mathbb{R}}^2$ and $\tau_{\mathbb{C}}$ implies that ϕ^{-1} is measurable relative to the Borel σ-algebra on $\mathbb{R}^2$ and $\mathcal{B}(\mathbb{C})$, it follows that f is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{C})$.

Now assume the converse, so that f is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{C})$. Then, h is measurable relative to $\mathcal{E}$ and the Borel σ-algebra on $\mathbb{R}^2$, since ϕ is continuous, and for any $A \in \mathbb{B}^1$,

$$(Re(f))^{-1}(A) = (Re(f))^{-1}(A) \cap (Im(f))^{-1}(\mathbb{R}) = h^{-1}(A \times \mathbb{R}) \in \mathcal{E}$$

because $A \times \mathbb{R}$ is an open rectangle and thus an element of the Borel σ-algebra on $\mathbb{R}^2$. It follows that $Re(f)$ is measurable relative to $\mathcal{E}$ and $\mathcal{B}(\mathbb{R})$, and by the same process, so is $Im(f)$.

Thus, a complex function f on E is $\mathcal{E}$-measurable if and only if its real and imaginary parts $Re(f)$ and $Im(f)$ are $\mathcal{E}$-measurable real functions.

Integrability of Complex Functions

The absolute value $|f|$ of f is defined as

$$|f| = \sqrt{Re(f)^2 + Im(f)^2}.$$

Since $|f|$ is a continuous function of the real valued functions $Re(f)$ and $Im(f)$, $|f|$ is a non-negative $\mathcal{E}$-measurable function.

A $\mathcal{E}$-measurable complex function f is said to be μ-integrable if $Re(f)$ and $Im(f)$ are μ-integrable real-valued functions. By the result we derived above, f is μ-integrable if and only if $\int_E |Re(f)|d\mu, \int_E |Im(f)|d\mu < +\infty$.
Suppose that $\int_E |f|d\mu < +\infty$. Then, because $|Re(f)|, |Im(f)| \leq |f|$, by the monotonicity of integration of non-negative functions,

$$\int_E |Re(f)|d\mu, \int_E |Im(f)|d\mu \leq \int_E |f|d\mu < +\infty$$

and f is μ-integrable.
Conversely, suppose that f is μ-integrable. Then, becauase $|f| \leq |Re(f)| + |Im(f)|$, by the monotonicity and linearity of integration of non-negative functions,

$$\int_E |f|d\mu \leq \int_E |Re(f)|d\mu + \int_E |Im(f)|d\mu < +\infty.$$

Therefore, as in the case of numerical functions, a complex valued $\mathcal{E}$-measurable function f is μ-integrable if and only if $|f|$ is μ-integrable, that is, $\int_E |f|d\mu < +\infty$.

We denote the set of all $\mathcal{E}$-measurable and μ-integrable complex-valued functions by $L^1(\mathcal{E}, \mu)$; this includes real valued functions that are $\mathcal{E}$-measurable and μ-integrable as well. In light of the above characterization, we can express $L^1(\mathcal{E}, \mu)$ as

$$L^1(\mathcal{E}, \mu) = \left\{ f : E \to \mathbb{C} \middle| Re(f), Im(f) \in \mathcal{E} \text{ and } \int_E |f|d\mu < +\infty \right\}.$$

The μ-integral for some function $f \in L^1(\mathcal{E}, \mu)$ over a set $A \in \mathcal{E}$ is defined as

$$\int_A f d\mu = \int_A Re(f)d\mu + i \cdot \int_A Im(f)d\mu,$$

where the integrals on the right are well-defined because $Re(f), Im(f)$ are μ-integrable real valued functions. It follows that $\int_A Re(f)d\mu$ and $\int_A Im(f)d\mu$ are the real and imaginary parts of the complex number $\int_A f d\mu$.

Integrals of complex-valued functions share much of the same properties as integrals of numerical functions:

Theorem 3.7 Let $(E, \mathcal{E}, \mu)$ be a measure space, and $f, g : E \to \mathbb{C}$ complex-valued functions that are $\mathcal{E}$-measurable. Then,

i) For any $f \in L^1(\mathcal{E}, \mu)$ and $A \in \mathcal{E}$,

$$\int_A f d\mu = \int_E (f \cdot I_A)\, d\mu.$$

ii) (Linearity) For any $z \in \mathbb{C}$, if $f, g \in L^1(\mathcal{E}, \mu)$, then $zf + g \in L^1(\mathcal{E}, \mu)$ and

$$\int_E (zf + g) d\mu = z \cdot \int_E f d\mu + \int_E g d\mu.$$

iii) If $f \in L^1(\mathcal{E}, \mu)$, then

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu.$$

Proof) i) Let $f \in L^1(\mathcal{E}, \mu)$ and $A \in \mathcal{E}$. Then, because $|fI_A| \leq |f|$ on E, by the monotonicity of integration

$$\int_E |fI_A| d\mu \leq \int_E |f| d\mu < +\infty$$

and $fI_A \in L^1(\mathcal{E}, \mu)$.
By definition, $Re(f)$ and $Im(f)$ are μ-integrable, and by the analogous result for integrable numerical functions,

$$\int_A Re(f) d\mu = \int_E (Re(f) \cdot I_A)\, d\mu \quad \text{and} \quad \int_A Im(f) d\mu = \int_E (Im(f) \cdot I_A)\, d\mu.$$

Furthermore, $Re(f)I_A = Re(fI_A)$ and $Im(f)I_A = Im(fI_A)$, so that

$$\begin{aligned} \int_A f d\mu &= \int_A Re(f) d\mu + i \cdot \int_A Im(f) d\mu \\ &= \int_E (Re(f)I_A) d\mu + i \cdot \int_E (Im(f)I_A) d\mu \\ &= \int_E Re(fI_A) d\mu + i \cdot \int_E Im(fI_A) d\mu \\ &= \int_E (f \cdot I_A) d\mu. \end{aligned}$$

Again, we will use this characterization from now on and focus on integrals over the entire set.

ii) Letting $a = Re(z)$ and $b = Im(z)$, note that

$$zf = (a + ib)(Re(f) + iIm(f)) = (a \cdot Re(f) - b \cdot Im(f)) + i(b \cdot Re(f) + a \cdot Im(f)),$$

so that

$$Re(zf) = a \cdot Re(f) - b \cdot Im(f) \quad \text{and} \quad Im(zf) = b \cdot Re(f) + a \cdot Im(f).$$

Since $Re(f), Im(f)$ are $\mathcal{E}$-measurable real-valued functions, $Re(zf)$ and $Im(zf)$ are also $\mathcal{E}$-measurable real-valued functions, implying that zf is $\mathcal{E}$-measurable. In addition,

$$zf + g = (Re(zf) + Re(g)) + i(Im(zf) + Im(g));$$

by implication, $Re(zf+g) = Re(zf) + Re(g)$, $Im(zf+g) = Im(zf) + Im(g)$, and because $Re(g)$ and $Im(g)$ are also $\mathcal{E}$-measurable real-valued functions, $Re(zf+g)$, $Im(zf+g)$ are $\mathcal{E}$-measurable real-valued functions, indicating that $zf+g$ is a $\mathcal{E}$-measurable complex function.

To show that $zf + g \in L^1(\mathcal{E}, \mu)$, note that

$$\begin{aligned} |zf+g| \le |Re(zf+g)| + |Im(zf+g)| &\le |Re(zf)| + |Re(g)| + |Im(zf)| + |Im(g)| \\ &\le (|a|+|b|)(|Re(f)| + |Im(f)|) + |Re(g)| + |Im(g)|. \end{aligned}$$

Since $Re(f), Im(f), Re(g), Im(g) \in L^1(\mathcal{E}, \mu)$ by the μ-integrability of f, g, by the monotonicity and linearity of integration of non-negative functions we have

$$\begin{aligned} \int_E |zf+g| d\mu \le (|a|+|b|) &\left(\int_E |Re(f)| d\mu + \int_E |Im(f)| d\mu \right) \\ &+ \left(\int_E |Re(g)| d\mu + \int_E |Im(g)| d\mu \right) < +\infty. \end{aligned}$$

This implies that $zf + g \in L^1(\mathcal{E}, \mu)$.

Finally, the integral of $zf + g$ can be decomposed as follows:

$$\begin{aligned} \int_E (zf+g) d\mu &= \int_E Re(zf+g) d\mu + i \cdot \int_E Im(zf+g) d\mu \\ &\hspace{6cm} \text{(Definition of the Integral)} \\ &= \int_E Re(zf) d\mu + \int_E Re(g) d\mu + i \cdot \int_E Im(zf) d\mu + i \cdot \int_E Im(g) d\mu \\ &\hspace{6cm} \text{(Linearity)} \\ &= a \cdot \int_E Re(f) d\mu + i^2 b \cdot \int_E Im(f) d\mu + ia \cdot \int_E Im(f) d\mu + ib \cdot \int_E Re(f) d\mu + \int_E g d\mu \\ &\hspace{6cm} \text{(Linearity)} \\ &= a \cdot \int_E f d\mu + ib \cdot \int_E f d\mu + \int_E g d\mu \\ &= z \cdot \int_E f d\mu + \int_E g d\mu. \end{aligned}$$

iii) If $\int_E f d\mu = 0$, then $|\int_E f d\mu| = 0 \leq \int_E |f| d\mu < +\infty$ trivially. Suppose now that $\int_E f d\mu \neq 0$.

We first prove that, for any nonzero complex number $z = a + ib \in \mathbb{C}$, there exists a complex number $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ such that $\alpha z = |z|$. Specifically, put $\alpha = \frac{\bar{z}}{|z|}$, which is well defined because $|z| > 0$; then,

$$|\alpha| = \frac{1}{|z|}|\bar{z}| = 1$$

and

$$\alpha z = \frac{z\bar{z}}{|z|} = \frac{|z|^2}{|z|} = |z|.$$

$\int_E f d\mu \in \mathbb{C}$ is a non-zero complex number, so by the above observation there exists a complex $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha \cdot \int_E f d\mu = |\int_E f d\mu|$. It follows that

$$\begin{aligned} \left|\int_E f d\mu\right| = \alpha \cdot \int_E f d\mu &= \int_E (\alpha f) d\mu \\ &= \int_E Re(\alpha f) d\mu && (\textstyle\int_E (\alpha f) d\mu \text{ is real-valued}) \\ &\leq \int_E |Re(\alpha f)| d\mu \\ &\leq \int_E |\alpha f| d\mu \\ &= |\alpha| \cdot \int_E |f| d\mu && (\text{Linearity}) \\ &= \int_E |f| d\mu. && (|\alpha| = 1) \end{aligned}$$

We can therefore see that, in any case,

$$\left|\int_E f d\mu\right| \leq \int_E |f| d\mu.$$

Q.E.D.

3.4 The Dominated Convergence Theorem

This section is dedicated to a workhorse theorem for interchanging limits and integrals. It utilizes Fatou's lemma to derive more general conditions under which limit taking and integrals can be interchanged.

Theorem 3.8 (The Dominated Convergence Theorem, DCT)
Let $(E,\mathcal{E},\mu)$ be a measure space, and $\{f_n\}_{n\in N_+}$ a sequence of functions in $L^1(\mathcal{E},\mu)$. Suppose that

- There exists a non-negative $\mathcal{E}$-measurable function g such that $|f_n| \leq g$ for any $n \in N_+$ and $\int_E g d\mu < +\infty$.
- $f_n \to f$ for some $\mathcal{E}$-measurable complex valued function f

Then, $f \in L^1(\mathcal{E},\mu)$ and

$$\lim_{n\to\infty} \int_E |f_n - f| d\mu = 0.$$

By implication,

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu \in \mathbb{C}.$$

Proof) Observe that $|f_n(x)| \leq g(x)$ for any $n \in N_+$ and $x \in E$. $Re(f_n) \to Re(f)$ and $Im(f_n) \to Im(f)$ as $n \to \infty$, and because measurability is preserved across limits, $Re(f), Im(f)$ are $\mathcal{E}$-measurable numerical functions, so that f is $\mathcal{E}$-measurable. Furthermore,

$$|f(x)| = \lim_{n\to\infty} |f_n(x)| \leq g(x)$$

for any $x \in E$, so by the monotonicity of integration,

$$\int_E |f| d\mu \leq \int_E g d\mu < +\infty,$$

so $f \in L^1(\mathcal{E},\mu)$.

We now define a sequence $\{g_n\}_{n\in N_+}$ of non-negative measurable functions: for any $n \in N_+$, define

$$g_n = 2g - |f_n - f|,$$

which is well-defined because $|f_n - f| \in [0,+\infty)$ for any $n \in N_+$ (all f_n and f are complex-valued). For any $x \in E$, since

$$|f_n(x) - f(x)| \leq |f_n(x)| + |f(x)| \leq 2g(x),$$

$g_n(x) = 2g(x) - |f_n(x) - f(x)| \in [0, +\infty]$. Therefore, $\{g_n\}_{n \in N_+}$ is a sequence of $\mathcal{E}$-measurable non-negative real-valued functions. In addition,

$$\int_E |f_n - f| d\mu \leq 2 \cdot \int_E g d\mu < +\infty$$

by the monotonicity and linearity of integration for any $n \in N_+$, so $\{|f_n - f|\}_{n \in N_+}$ is a sequence of μ-integrable non-negative functions.

By Fatou's lemma,

$$\int_E \left(\liminf_{n \to \infty} g_n \right) d\mu \leq \liminf_{n \to \infty} \left(\int_E g_n d\mu \right).$$

We will now study both sides of the above inequality:

1) **The left hand side**
   Note that

$$\liminf_{n \to \infty} g_n = 2g - \left(\limsup_{n \to \infty} |f_n - f| \right);$$

   because $\limsup_{n \to \infty} |f_n - f| = \lim_{n \to \infty} |f_n - f| = 0$ by assumption,

$$\int_E \left(\liminf_{n \to \infty} g_n \right) d\mu = 2 \cdot \int_E g d\mu.$$

2) **The right hand side**
   For any $n \in N_+$,

$$\int_E g_n d\mu = 2 \cdot \int_E g d\mu - \int_E |f_n - f| d\mu$$

   by the linearity of integration. Thus,

$$\liminf_{n \to \infty} \left(\int_E g_n d\mu \right) = 2 \cdot \int_E g d\mu - \limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right).$$

Therefore, the above inequality can be rewritten as

$$2 \cdot \int_E g d\mu \leq 2 \cdot \int_E g d\mu - \limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right).$$

$\int_E g d\mu \in [0, +\infty)$ by assumption, so

$$\limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right) \leq 0.$$

For any $n \in N_+$, $|f_n - f|$ is a non-negative valued measurable function, so $\int_E |f_n - f| d\mu$

is non-negative valued and thus

$$0 \leq \liminf_{n\to\infty} \left(\int_E |f_n - f| d\mu \right).$$

Putting the two inequalities together, it holds that

$$0 \leq \liminf_{n\to\infty} \left(\int_E |f_n - f| d\mu \right) \leq \limsup_{n\to\infty} \left(\int_E |f_n - f| d\mu \right) \leq 0,$$

or

$$\liminf_{n\to\infty} \left(\int_E |f_n - f| d\mu \right) = \limsup_{n\to\infty} \left(\int_E |f_n - f| d\mu \right) = 0.$$

Therefore,

$$\lim_{n\to\infty} |f_n - f| d\mu = 0.$$

$f_n, f \in L^1(\mathcal{E}, \mu)$ for any $n \in N_+$, so by the linearity of integration,

$$\int_E (f_n - f) d\mu = \int_E f_n d\mu - \int_E f d\mu.$$

$\int_E f d\mu \in \mathbb{C}$, and by theorem 3.7,

$$\left| \int_E f_n d\mu - \int_E f d\mu \right| = \left| \int_E (f_n - f) d\mu \right| \leq \int_E |f_n - f| d\mu.$$

Taking $n \to \infty$ on both sides reveals that

$$\lim_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu.$$

Q.E.D.

We now present important corollaries of the DCT, like we did for the MCT.

3.4.1 The Bounded Convergence Theorem

The first corollary we present is a simplified version of the DCT when the associated measure is finite. This result is often used in probability theory to interchange limits and expectations, since probability measures are finite by definition. An example includes the proof that the derivatives of the moment generating function yield the moments of a random variable. The formal statement and proof are as follows:

Corollary to Theorem 3.8 (Bounded Convergence Theorem, BCT)
Let $(E, \mathcal{E}, \mu)$ be a finite measure space, and $\{f_n\}_{n \in N_+}$ a sequence of $\mathcal{E}$-measurable bounded complex functions that converges poinwise to some $\mathcal{E}$-measurable complex function f.
Then, each f_n and f are μ-integrable, and

$$\lim_{n \to \infty} \int_E |f_n - f| d\mu = 0,$$

which implies that $\int_E f d\mu = \lim_{n \to \infty} \int_E f_n d\mu$.

Proof) Suppose that there exists an $M > 0$ such that $|f_n| \leq M$ for any $n \in N_+$. Then, defining $g : E \to [0, +\infty]$ as $g(x) = M$ for any $x \in E$, g is trivially measurable, and

$$\int_E g d\mu = M \cdot \mu(E) < +\infty,$$

where the finiteness of μ was used ($\mu(E) < +\infty$). Therefore, $\{f_n\}_{n \in N_+}$ is a sequence dominated by an integrable function g that converges to some complex measurable function f. As such, by the DCT, $f \in L^1(\mathcal{E}, \mu)$ and

$$\int_E f d\mu = \lim_{n \to \infty} \int_E f_n d\mu.$$

Q.E.D.

The fact that the condition that μ is a finite measure is necessary for the above result can be seen through the following simple counterexample.
Consider the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, where λ is the Lebesgue measure on $\mathbb{R}$. Clearly, λ is not a finite measure (we showed in chapter 2 that it is instead σ-finite).
Now suppose the sequence $\{f_n\}_{n \in N_+}$ of functions on $\mathbb{R}$ are defined as $f_n(x) = \frac{1}{n}$ for any $x \in \mathbb{R}$ and $n \in N_+$. Then, $\{f_n\}_{n \in N_+}$ is bounded above by 1 and converges to the measurable function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 0$ for any $x \in E$. However,

$$\int_{\mathbb{R}} f d\lambda = 0 \cdot \lambda(\mathbb{R}) = 0, \quad \text{while} \quad \int_{\mathbb{R}} f_n d\lambda = \frac{1}{n} \cdot \lambda(\mathbb{R}) = +\infty \quad \text{for any } n \in N_+.$$

Therefore,

$$\int_{\mathbb{R}} f d\lambda = 0 \neq +\infty = \lim_{n \to \infty} \int_{\mathbb{R}} f_n d\lambda.$$

3.4.2 The Generalized DCT

The next corollary is a generalized version of the DCT, in which the dominating function is replaced by a dominating sequence that converges pointwise and integral-wise to some function. In this case, the dominating sequence is usually taken to be an increasing sequence of functions so that the MCT can be applied to it. The formal statement and proof are as follows:

Corollary to Theorem 3.8 (The Generalized DCT)
Let $(E, \mathcal{E}, \mu)$ be a measure space, and $\{f_n\}_{n\in N_+}$ a sequence in $L^1(\mathcal{E}, \mu)$ and $\{g_n\}_{n\in N_+}$ a sequence in $\mathcal{E}_+$ such that:

- For any $n \in N_+$, $|f_n| \leq g_n$.
- There exists a $g \in \mathcal{E}_+$ such that $g_n \to g$ pointwise and

$$\lim_{n\to\infty} \int_E g_n d\mu = \int_E g d\mu < +\infty.$$

- There exists a complex measurable f such that $f_n \to f$ pointwise.

Then, $f \in L^1(\mathcal{E}, \mu)$ and

$$\lim_{n\to\infty} \int_E |f_n - f| d\mu = 0,$$

which implies that

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu.$$

Proof) The proof proceeds almost beat by beat like the proof of the DCT.
Since $|f_n| \leq g_n$ for any $n \in N_+$ and $f_n \to f$, $g_n \to g$ pointwise, we have $|f| \leq g$. By the monotonicity of integration,

$$\int_E |f| d\mu \leq \int_E g d\mu < +\infty,$$

so $f \in L^1(\mathcal{E}, \mu)$.

For any $n \in N_+$, define the sequence $\{h_n\}_{n\in N_+}$ as

$$h_n = g_n + g - |f_n - f|,$$

which is well defined because $|f_n - f|$ is always complex-valued. In addition, for any $x \in E$,

$$|f_n(x) - f(x)| \leq |f_n(x)| + |f(x)| \leq g_n(x) + g(x),$$

so that $h_n(x) = g_n(x) + g(x) - |f_n(x) - f(x)| \in [0, +\infty]$. Therefore, $\{h_n\}_{n \in N_+}$ is a sequence of non-negative measurable functions. We can also see that

$$\int_E |f_n - f| d\mu \leq \int_E g_n d\mu + \int_E g d\mu$$

by the monotonicity and linearity of integration. $\{\int_E g_n d\mu\}_{n \in N_+}$ is a sequence in $[0, +\infty]$ that converges to the real number $\int_E g d\mu$, so it is a bounded sequence, which implies that

$$\int_E |f_n - f| d\mu \leq \int_E g_n d\mu + \int_E g d\mu < +\infty.$$

This means that $\{g_n\}_{n \in N_+}$ and $\{|f_n - f|\}_{n \in N_+}$ are sequences of μ-integrable non-negative functions.

By Fatou's lemma,

$$\int_E \left(\liminf_{n \to \infty} h_n \right) d\mu \leq \liminf_{n \to \infty} \left(\int_E h_n d\mu \right).$$

Since $\lim_{n \to \infty} g_n = g$ and $\lim_{n \to \infty} |f_n - f| = 0$, we have

$$\liminf_{n \to \infty} h_n = \lim_{n \to \infty} h_n = 2g,$$

and as such

$$\int_E \left(\liminf_{n \to \infty} h_n \right) d\mu = 2 \cdot \int_E g d\mu$$

by the linearity of integration.
On the other hand,

$$\begin{aligned}
\liminf_{n \to \infty} \left(\int_E h_n d\mu \right) &= \liminf_{n \to \infty} \left(\int_E g d\mu + \int_E g_n d\mu - \int_E |f_n - f| d\mu \right) \\
&\qquad \text{(Linearity of Integration)} \\
&= 2 \cdot \int_E g d\mu - \limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right). \\
&\qquad \left(\textstyle\int_E g_n d\mu \to \int_E g d\mu \text{ by assumption}\right)
\end{aligned}$$

Therefore, the above inequality reduces to

$$2 \cdot \int_E g d\mu \leq 2 \cdot \int_E g d\mu - \limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right),$$

and because $2 \cdot \int_E g d\mu < +\infty$, we have

$$\limsup_{n \to \infty} \left(\int_E |f_n - f| d\mu \right) \leq 0.$$

By implication,

$$\limsup_{n\to\infty}\left(\int_E |f_n - f| d\mu\right) = \liminf_{n\to\infty}\left(\int_E |f_n - f| d\mu\right) = \lim_{n\to\infty}\left(\int_E |f_n - f| d\mu\right) = 0,$$

which also tells us that

$$\lim_{n\to\infty}\int_E f_n d\mu = \int_E f d\mu.$$

Q.E.D.

3.4.3 Scheffe's Lemma

This result concerns the equivalence of L^1 convergence and convergence in mean for a sequence of measurable functions that converges pointwise. It is used in probability theory to establish that the convergence of density functions implies convergence in distribution.

Corollary to Theorem 3.8 (Scheffe's Lemma)
Let $(E, \mathcal{E}, \mu)$ be a finite measure space, and $\{f_n\}_{n \in N_+}$ a sequence in $L^1(\mathcal{E}, \mu)$ that converges poinwise to some $\mathcal{E}$-measurable complex function f. Then,

$$\lim_{n \to \infty} \int_E |f_n - f| d\mu = 0$$

if and only if

$$\lim_{n \to \infty} \int_E |f_n| d\mu = \int_E |f| d\mu,$$

and $f \in L^1(\mathcal{E}, \mu)$ under either condition.

Proof) Suppose initially that

$$\lim_{n \to \infty} \int_E |f_n - f| d\mu = 0.$$

Then, $\{\int_E |f_n - f| d\mu\}_{n \in N_+}$ is a convergent and thus bounded sequence in $[0, +\infty]$. By implication, $\{|f_n - f|\}_{n \in N_+}$ is a sequence of non-negative functions in $L^1(\mathcal{E}, \mu)$, and because

$$|f| \leq |f_1| + |f_1 - f|,$$

by the monotonicity and linearity of integration

$$\int_E |f| d\mu \leq \int_E |f_1| d\mu + \int_E |f_1 - f| d\mu < +\infty.$$

We have seen that $f \in L^1(\mathcal{E}, \mu)$.
In addition, by the inequality in theorem 3.7,

$$\begin{aligned}
\left| \int_E |f_n| d\mu - \int_E |f| d\mu \right| &= \left| \int_E (|f_n| - |f|) \, d\mu \right| && \text{(Linearity of Integration)} \\
&\leq \int_E ||f_n| - |f|| d\mu \\
&\leq \int_E |f_n - f| d\mu && \text{(The triangle inequality)}
\end{aligned}$$

for any $n \in N_+$. Therefore,

$$\lim_{n \to \infty} \left| \int_E |f_n| d\mu - \int_E |f| d\mu \right| = 0$$

and

$$\lim_{n\to\infty}\int_E |f_n| d\mu = \int_E |f| d\mu.$$

Conversely, suppose that

$$\lim_{n\to\infty}\int_E |f_n| d\mu = \int_E |f| d\mu.$$

Then, defining $|f_n| = g_n$ for any $n \in N_+$, $|f_n| \le g_n$ for any $n \in N_+$, where $\{g_n\}_{n\in N_+} \subset \mathcal{E}_+$ satisfies $g_n \to g = |f|$ pointwise and

$$\int_E g_n d\mu = \int_E |f_n| d\mu \to \int_E |f| d\mu = \int_E g d\mu$$

as $n \to \infty$. By the generalized DCT, then, $f \in L^1(\mathcal{E},\mu)$ and

$$\lim_{n\to\infty}\int_E |f_n - f| d\mu = 0.$$

Q.E.D.

To see that this result does lead to the conclusion that convergence of densities implies convergence in distribution, let $\{\mu_n\}_{n\in N_+}$ be a sequence of distributions on $(E,\mathcal{E})$ and μ a distribution in $(E,\mathcal{E})$. Suppose λ is some σ-finite measure with respect to which μ_n, μ are absolutely continuous and thus, by the Radon-Nikodym theorem, with respect to which μ_n, μ have densities $f_n, f \in \mathcal{E}_+$ for any $n \in N_+$.
Now suppose that $f_n \to f$ pointwise. Then, because

$$\lim_{n\to\infty}\int_E f_n d\mu = 1 = \int_E f d\mu,$$

by Scheffe's lemma

$$\lim_{n\to\infty}\int_E |f_n - f| d\mu = 0.$$

Therefore, for any $A \in \mathcal{E}$, because

$$\left|\int_A f_n d\mu - \int_A f d\mu\right| = \left|\int_E (f_n - f)\cdot I_A d\mu\right| \le \int_E |f_n - f| d\mu$$

for any $n \in N_+$, the fact that $\lim_{n\to\infty}\int_E |f_n - f| d\mu = 0$ implies that

$$\lim_{n\to\infty}\left|\int_A f_n d\mu - \int_A f d\mu\right| = 0,$$

or equivalently,

$$\lim_{n\to\infty}\int_A f_n d\mu = \int_A f d\mu.$$

By the definition of f_n, f as densities, this means that

$$\lim_{n\to\infty}\mu_n(A) = \mu(A)$$

for any $A \in \mathcal{E}$. By the portmanteau theorem, this implies that $\mu_n \to \mu$ weakly, or in distribution. In fact, it is possible to show the stronger result that $\mu_n \to \mu$ in total variation norm by observing that

$$\sup_{A\in\mathcal{E}}\left|\int_A f_n d\mu - \int_A f d\mu\right| \leq \int_E |f_n - f| d\mu,$$

which would imply that

$$\lim_{n\to\infty}\sup_{A\in\mathcal{E}}|\mu_n(A) - \mu(A)| = 0,$$

which basically means that the probability of any event under μ_n converges uniformly to that probability under μ.

3.5 Almost Everywhere

3.5.1 Definition and Properties that hold a.e.

Let $(E, \mathcal{E}, \mu)$ be a measure space. We say that some property holds almost everywhere on E if it holds true for any $x \in A$ such that $\mu(A^c) = 0$; since this is clearly measure-dependent, we denote this by "a.e. $[\mu]$".
Suppose f, g are $\mathcal{E}$-measurable numerical or complex functions. By definition, f and g are equal almost everywhere on E, or $f = g$ a.e. on $[\mu]$, if

$$\mu(\{f \neq g\}) = 0.$$

Note that $\{f \neq g\}$ is a measurable set because both f and g are measurable. Such almost everywhere equivalence implies that their integrals over any measurable set is equal, as shown in the next theorem:

Theorem 3.9 Let $(E, \mathcal{E}, \mu)$ be a measure space, and f, g numerical or complex functions that are $\mathcal{E}$-measurable. If $f = g$ a.e. $[\mu]$ and f, g are non-negative functions, then

$$\int_E f d\mu = \int_E g d\mu.$$

If $f = g$ a.e. $[\mu]$ and f is μ-integrable, then so is g, and

$$\int_E f d\mu = \int_E g d\mu.$$

Proof) Suppose initially that $f, g \in \mathcal{E}_+$. Then,

$$\begin{aligned}
\int_E f d\mu &= \int_{\{f=g\}} f d\mu + \int_{\{f \neq g\}} f d\mu \\
&= \int_{\{f=g\}} f d\mu && \text{(theorem 3.2)} \\
&= \int_{\{f=g\}} g d\mu && (f = g \text{ on } \{f = g\}) \\
&= \int_E g d\mu. && \text{(Same process as above)}
\end{aligned}$$

Now let f be a μ-integrable numerical function. Suppose $f(x) = g(x)$ for some $x \in E$. If $f(x) = g(x) \geq 0$, then

$$f^+(x) = f(x) = g(x) = g^+(x)$$

and

$$f^-(x) = 0 = g^-(x),$$

while if $f(x) = g(x) < 0$, then

$$f^-(x) = f(x) = g(x) = g^-(x)$$

and

$$f^+(x) = 0 = g^+(x).$$

Therefore, $\{f = g\} \subset \{f^+ = g^+\} \cap \{f^- \cap g^-\}$. Conversely, if $f^\pm(x) = g^\pm(x)$ for some $x \in E$, then

$$f(x) = f^+(x) - f^-(x) = g^+(x) - g^-(x) = g(x),$$

so that $\{f^+ = g^+\} \cap \{f^- \cap g^-\} \subset \{f = g\}$ and thus $\{f = g\} = \{f^+ = g^+\} \cap \{f^- \cap g^-\}$. It follows that

$$\{f \neq g\} = \{f^+ \neq g^+\} \cup \{f^- \neq g^-\},$$

and by the monotonicity of measures, $\mu(\{f \neq g\}) = 0$ implies

$$\mu(\{f^\pm \neq g^\pm\}) = 0.$$

By the result shown above, we can now conclude that

$$\int_E f^\pm d\mu = \int_E g^\pm d\mu < +\infty,$$

so that g is also μ-integrable. Furthermore, by the definition of the integral for integrable functions,

$$\begin{aligned}\int_E f d\mu &= \int_E f^+ d\mu - \int_E f^- d\mu \\ &= \int_E g^+ d\mu - \int_E g^- d\mu = \int_E g d\mu.\end{aligned}$$

Finally, let $f \in L^1(\mathcal{E}, \mu)$. Then, it is clear that $\{f = g\} = \{Re(f) = Re(g)\} \cap \{Im(f) = Im(g)\}$, where $Re(f), Im(f), Re(g), Im(g)$ are all μ-integrable real valued functions. As such,

$$\{f \neq g\} = \{Re(f) \neq Re(g)\} \cup \{Im(f) \neq Im(g)\},$$

and by the monotonicity of measures,

$$\mu(\{Re(f) \neq Re(g)\}) = \mu(\{Im(f) \neq Im(g)\}) = 0,$$

so that $Re(g), Im(g)$ are μ-integrable and

$$\int_E Re(f)d\mu = \int_E Re(g)d\mu \quad \text{and} \quad \int_E Im(f)d\mu = \int_E Im(g)d\mu$$

by the results shown above. This implies that $g \in L^1(\mathcal{E}, \mu)$ and

$$\begin{aligned}\int_E f d\mu &= \int_E Re(f)d\mu + i \cdot \int_E Im(f)d\mu \\ &= \int_E Re(g)d\mu + i \cdot \int_E Im(f)d\mu.\end{aligned}$$

Q.E.D.

We have shown that, no matter the target space of f, g, if they are equal μ-a.e., then their μ-integrals are equal over any set $A \in \mathcal{E}$, provided that they are their integrals are well-defined. This shows us that, insofar as integration is concerned, we might be able to identify functions that are equal almost everywhere.

Recall the vanshing and finiteness properties (theorem 3.3). Those properties can now be stated as follows:

- **Vanishing Property**
  For any $f \in \mathcal{E}_+$ and $A \in \mathcal{E}$,
  $$\int_A f d\mu = 0$$
  if and only if $f = 0$ μ-almost everywhere on A.

- **Finiteness Proprety**
  For any $f \in \mathcal{E}_+$ and $A \in \mathcal{E}$, if
  $$\int_A f d\mu < +\infty,$$
  then $f < +\infty$ μ-almost everywhere on A.

The above results were stated only for non-negative functions. Now, we make use of the almost everywhere notation and prove similar results for arbitrary complex measurable functions as well.

Theorem 3.10 Let $(E, \mathcal{E}, \mu)$ be a measure space. Then, the following hold true:

i) For any $f \in L^1(\mathcal{E}, \mu)$, if
$$\int_A f d\mu = 0$$
for any $A \in \mathcal{E}$, then $f = 0$ a.e. $[\mu]$.

ii) For any $f \in L^1(\mathcal{E}, \mu)$, if
$$\left| \int_E f d\mu \right| = \int_E |f| d\mu,$$
then there exists an $\alpha \in \mathbb{C}$ such that $\alpha f = |f|$ a.e. $[\mu]$.

iii) For any sequence $\{A_n\}_{n \in N_+}$ of measurable sets such that
$$\sum_{n=1}^{\infty} \mu(A_n) < +\infty,$$
μ-almost every $x \in E$ is contained in at most a finite number of sets in $\{A_n\}_{n \in N_+}$.

Proof) i) Initially assume that f is real-valued. By assumption, for any $A \in \mathcal{E}$

$$\int_A f d\mu = \int_A f^+ d\mu - \int_A f^- d\mu = 0,$$

or

$$\int_A f^+ d\mu = \int_A f^- d\mu,$$

which follows because $\int_A f^\pm d\mu < +\infty$ by integrability.
Define the set $H = \{x \in E \mid f(x) = 0\}$. Then,

$$H^c = \{x \in E \mid f^+(x) \neq 0\} \cup \{x \in E \mid f^-(x) \neq 0\} = A^+ \cup A^-.$$

It suffices to show that $\mu(A^\pm) = 0$.
Suppose that $\mu(A^+) > 0$. Then, noting that

$$\begin{aligned} A^+ &= \{f^+ \neq 0\} = \{f^+ > 0\} && (f^+ \text{ is a non-negative function}) \\ &= \bigcup_n \{f^+ > \frac{1}{n}\} = \bigcup_n A_n^+, \end{aligned}$$

by subadditivity

$$0 < \mu(A^+) \leq \sum_{n=1}^{\infty} \mu(A_n^+),$$

so that there must exist some $n \in N_+$ such that $\mu(A_n^+) > 0$ (otherwise, $\mu(A^+) = 0$, a contradiction). It follows that

$$\int_{A_n^+} f^+ d\mu \geq \frac{1}{n} \mu(A_n^+) > 0.$$

However, because $f^-(x) = 0$ for any $x \in A_n^+$ ($\min(f(x), 0) = 0$ if $f(x) > 0$), $f^-(x) = 0$ for any $x \in A_n^+$, which implies that

$$0 = \int_{A_n^+} f^- d\mu = \int_{A_n^+} f^+ d\mu > 0,$$

a contradiction. Therefore, $\mu(A^+) = 0$.
Through a similar process, we can show that $\mu(A^-) = 0$. Therefore,

$$\mu(H^c) \leq \mu(A^+) + \mu(A^-) = 0$$

and $\mu(H^c) = 0$, so that $f = 0$ a.e. $[\mu]$.

Now let $f \in L^1(\mathcal{E},\mu)$ in general. Then, for any $A \in \mathcal{E}$,

$$\int_A f d\mu = \int_A Re(f) d\mu + i \cdot \int_A Im(f) d\mu = 0,$$

so that

$$\int_A Re(f) d\mu = \int_A Im(f) d\mu = 0.$$

By the previous result, there exist sets $A, B \in \mathcal{E}$ such that $A = \{Re(f) = 0\}$, $B = \{Im(f) = 0\}$ and $\mu(A^c) = \mu(B^c) = 0$. Since $\{f = 0\} = A \cap B$,

$$\mu(\{f \neq 0\}) \leq \mu(A^c) + \mu(B^c) = 0,$$

or $\mu(\{f \neq 0\}) = 0$. Therefore, $f = 0$ a.e. $[\mu]$.

ii) Suppose that

$$\left|\int_E f d\mu\right| = \int_E |f| d\mu$$

for some $f \in L^1(\mathcal{E},\mu)$. If $\int_E f d\mu = 0$, then by the vanishing property for non-negative functions $|f| = 0$, or $f = 0$, a.e. $[\mu]$, indicating that $0 \cdot f = |f|$ a.e. $[\mu]$. Now suppose that $\int_E f d\mu \neq 0$. Then, as noted in theorem 3.7, defining

$$\alpha = \frac{\overline{\int_E f d\mu}}{|\int_E f d\mu|} \in \mathbb{C},$$

we have $|\alpha| = 1$ and

$$\alpha \cdot \left(\int_E f d\mu\right) = \left|\int_E f d\mu\right|.$$

By the linearity of integration, we now have

$$\int_E (\alpha f) d\mu = \left|\int_E f d\mu\right| = \int_E |f| d\mu,$$

where the latter equality follows by assumption. Since αf must be real valued and

$$\alpha f \leq |\alpha f| = |\alpha||f| = |f|,$$

by the linearity of integration

$$\int_E (|f| - \alpha f)\, d\mu = 0,$$

where $|f| - \alpha f$ is a non-negative measurable function. Finally, by the vanishing property of non-negative functions,

$$|f| - \alpha f = 0,$$

or $|f| = \alpha f$, a.e. $[\mu]$.

iii) Note that $\{I_{A_n}\}_{n \in N_+}$ is a sequence of non-negative measurable functions; by the MCT for series, the non-negative function f defined as

$$f(x) = \sum_{n=1}^{\infty} I_{A_n}(x)$$

for any $x \in E$ is measurable, and

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_E I_{A_n} d\mu = \sum_{n=1}^{\infty} \mu(A_n) < +\infty,$$

where the last inequality follows by assumption.
Therefore, by the finiteness property of non-negative functions, $f < +\infty$ a.e. $[\mu]$.
If some $x \in E$ is contained in infinitely many of the A_n, then $f(x) = +\infty$, so it follows that almost every $x \in E$ is contained in finitely many of the A_n.

Q.E.D.

3.5.2 Complete Measure Spaces

The above finding indicates that we can extend our definition of measurability to encompass functions that are defined on the entire set E except for a set of measure 0. Formally, let $A \in \mathcal{E}$ be a set such that $\mu(A^c) = 0$, and let $f : A \to \mathbb{C}$ be a function. We will say that f is $\mathcal{E}$-measurable if, for any $B \in \mathcal{B}(\mathbb{R})$,

$$(Re(f))^{-1}(B) \cap A, (Im(f))^{-1}(B) \cap A \in \mathcal{E}.$$

Suppose that $f : A \to \mathbb{C}$ is $\mathcal{E}$-measurable, and define its trivial extension $\bar{f} : E \to \mathbb{C}$ as

$$\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in A \\ 0 & \text{if } x \in A^c. \end{cases}$$

Then, $A = \{f = \bar{f}\} = \{Re(f) = Re(\bar{f})\} \cap \{Im(f) = Im(\bar{f})\}$, so that, for any $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} (Re(\bar{f}))^{-1}(B) &= \Big((Re(\bar{f}))^{-1}(B) \cap A\Big) \cup \Big((Re(\bar{f}))^{-1}(B) \cap A^c\Big) \\ &= \begin{cases} ((Re(f))^{-1}(B) \cap A) \cup A^c & \text{if } 0 \in B \\ (Re(f))^{-1}(B) \cap A & \text{if } 0 \notin B \end{cases}. \end{aligned}$$

Because $(Re(f))^{-1}(B) \cap A \in \mathcal{E}$ by assumption and $A, A^c \in \mathcal{E}$, $Re(\bar{f})$ is an $\mathcal{E}$-measurable real-valued function. Likewise, $Im(\bar{f})$ is $\mathcal{E}$-measurable, indiciating that $\bar{f}$ is also measurable.
Thus, the definition of measurability for f given above is really nothing more than requiring the trivial extension $\bar{f}$ of f to the entire set E to be a measurable function.
Given a $\mathcal{E}$-measurable function $f : A \to \mathbb{C}$, then, we say that f is μ-integrable if its trivial extension $\bar{f}$ is μ-integrable, and define its integral as

$$\int_E f d\mu = \int_E \bar{f} d\mu.$$

However, we cannot yet consider any extension of f to E a measurable function aside from the trivial extension. To see this, consider an arbitrary extension g of f to E. Then, for any $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} (Re(g))^{-1}(B) &= \Big((Re(g))^{-1}(B) \cap A\Big) \cup \Big((Re(g))^{-1}(B) \cap A^c\Big) \\ &= \Big((Re(f))^{-1}(B) \cap A\Big) \cup \Big((Re(g))^{-1}(B) \cap A^c\Big). \end{aligned}$$

The first term is $\mathcal{E}$-measurable, but the second term, despite being a subset of a measure zero set A^c, is not guaranteed to be in the σ-algebra $\mathcal{E}$ and thus we cannot say that $(Re(g))^{-1}(B)$ is in $\mathcal{E}$. This means that g is not necessarily $\mathcal{E}$-measurable.

More generally, consider a $\mathcal{E}$-measurable real function f and a function $g : E \to \mathbb{R}$ such that

$f = g$ a.e. $[\mu]$, or there exists an $A \in \mathcal{E}$ such that $f(x) = g(x)$ for any $x \in A$ and $\mu(A^c) = 0$. The fact that $f = g$ a.e. $[\mu]$ does not guarantee that g is also $\mathcal{E}$-measurable; this can be seen from the fact that, for any $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned} g^{-1}(B) &= \left(g^{-1}(B) \cap A\right) \cup \left(g^{-1}(B) \cap A^c\right) \\ &= \left(f^{-1}(B) \cap A\right) \cup \left(g^{-1}(B) \cap A^c\right), \end{aligned}$$

since $f = g$ on A. By the measurability of f, $f^{-1}(B) \cap A \in \mathcal{E}$, but while $g^{-1}(B) \cap A^c$ is a subset of the measure zero set A^c, we do not know at the moment whether $g^{-1}(B) \cap A^c$ is a measurable set. Therefore, we cannot say for certain that g is $\mathcal{E}$-measurable.

The above problem will be immediately solved if any subset of a measure zero set, called negligible sets, are mesurable and of measure zero as well. Such measure spaces are called complete. Fortunately, for any arbitrary measure space $(E, \mathcal{E}, \mu)$ we are able to extend $\mathcal{E}$ and μ in a manner that makes it complete. This is shown below:

Theorem 3.11 (The Completion of Measure Spaces)
Let $(E,\mathcal{E},\mu)$ be a measure space, and define $\mathcal{N}$ as the set of all negligible sets on E, that is,

$$\mathcal{N} = \{A \subset E \mid \exists B \in \mathcal{E} \text{ s.t. } A \subset B, \mu(B) = 0\}.$$

Now define the collection of sets

$$\bar{\mathcal{E}} = \{A \subset E \mid \exists B, C \in \mathcal{E} \text{ s.t. } C \subset A \subset B, \mu(B \setminus C) = 0\}.$$

Then, $\bar{\mathcal{E}}$ is a σ-algebra on E.
Moreover, defining the function $\bar{\mu} : \bar{\mathcal{E}} \to [0,+\infty]$ as

$$\bar{\mu}(A) = \mu(C)$$

for any $A \in \bar{\mathcal{E}}$, where $B, C \in \mathcal{E}$ satisfy $C \subset A \subset B$ and $\mu(B \setminus C) = 0$, $\bar{\mu}$ is a measure on $(E,\bar{\mathcal{E}})$ such that $\bar{\mu}(A) = \mu(A)$ for any $A \in \mathcal{E}$.
Finally,

$$\bar{\mathcal{E}} = \mathcal{E} \cup \mathcal{N} = \{A \cup B \mid A \in \mathcal{E}, B \in \mathcal{N}\},$$

and for any $A \in \bar{\mathcal{E}} = \mathcal{E} \cup \mathcal{N}$ such that $A = B \cup N$ for some $B \in \mathcal{E}$ and $N \in \mathcal{N}$,

$$\bar{\mu}(A) = \mu(B).$$

Proof) We first show that $\bar{\mathcal{E}}$ is a σ-algebra on E.

i) Because $E \subset E \subset E$ and $\mu(E \setminus E) = 0$, where $E \in \mathcal{E}$, $E \in \bar{\mathcal{E}}$ by definition.

ii) For any $A \in \bar{\mathcal{E}}$, let $B, C \in \mathcal{E}$ satisfy $C \subset A \subset B$ and $\mu(B \setminus C) = 0$. Then, $B^c \subset A^c \subset C^c$ for $B^c, C^c \in \mathcal{E}$ and

$$\mu(C^c \setminus B^c) = \mu(C^c \cap B) = \mu(B \setminus C) = 0,$$

so by definition $A^c \in \bar{\mathcal{E}}$.

iii) For any sequence $\{A_n\}_{n \in N_+} \subset \bar{\mathcal{E}}$ with union $A = \bigcup_n A_n$, let $\{C_n\}_{n \in N_+}, \{B_n\}_{n \in N_+} \subset \mathcal{E}$ be sequences such that $C_n \subset A_n \subset B_n$ and $\mu(B_n \setminus C_n) = 0$ for any $n \in N_+$. Then, defining $C = \bigcup_n C_n$ and $B = \bigcup_n B_n$, $B, C \in \mathcal{E}$, and for any $n \in N_+$,

$$C \subset A = \bigcup_n A_n \subset B.$$

Furthermore,

$$\begin{aligned}
\mu(B \setminus C) = \mu\left(\left(\bigcup_n B_n\right) \cap C^c\right) &= \mu\left(\bigcup_n (B_n \cap C^c)\right) \\
&\leq \sum_{n=1}^{\infty} \mu(B_n \setminus C) && \text{(Countable Subadditivity)} \\
&\leq \sum_{n=1}^{\infty} \mu(B_n \setminus C_n). && \text{(Monotonicity; } C_n \subset C \text{ for any } n\text{)}
\end{aligned}$$

Since $\mu(B_n \setminus C_n) = 0$ for any $n \in N_+$, we have

$$\mu(B \setminus C) = 0,$$

and $A \in \bar{\mathcal{E}}$ by definition.

Now let the set function $\bar{\mu}$ be defined as above. We first check that $\bar{\mu}$ is actually a function. For any $A \in \bar{\mathcal{E}}$, suppose that there exist two pairs of sets, B_1, C_1 and B_2, C_2, in $\mathcal{E}$ such that $C_i \subset A \subset B_i$ and $\mu(B_i \setminus C_i) = 0$ for $i = 1, 2$. Then,

$$\mu(B_i) = \mu(B_i \setminus C_i) + \mu(C_i) = \mu(C_i)$$

for $i = 1, 2$ by finite additivity. Furthermore,

$$\begin{aligned}
\mu(B_1) &= \mu(B_1 \setminus B_2) + \mu(B_2) && \text{(Finite Additivity)} \\
&\leq \mu(B_1 \setminus C_1) + \mu(B_2) && \text{(Monotonicity; } C_1 \subset A \subset B_2\text{)} \\
&= \mu(B_2),
\end{aligned}$$

so that $\mu(B_1) \leq \mu(B_2)$. This holds the opposite direction as well, which implies that $\mu(B_2) \leq \mu(B_1)$ and thus $\mu(B_1) = \mu(B_2)$. As such, we have

$$\bar{\mu}(A) = \mu(C_1) = \mu(B_1) = \mu(B_2) = \mu(C_2),$$

so that $\bar{\mu}(A)$ is uniquely defined.
For any $A \in \mathcal{E}$, this indicates that $\bar{\mu}(A) = \mu(A)$.

It is now easy to show that $\bar{\mu}$ is a measure on $(E, \bar{\mathcal{E}})$:

i) $\bar{\mu}(\emptyset) = \mu(\emptyset) = 0$ because $\emptyset \in \mathcal{E}$.

ii) For any disjoint sequence $\{A_n\}_{n \in N_+} \subset \bar{\mathcal{E}}$ with union $A = \bigcup_n A_n$, recall there exist sequences $\{C_n\}_{n \in N_+}, \{B_n\}_{n \in N_+} \subset \mathcal{E}$ be sequences such that $C_n \subset A_n \subset B_n$ and $\mu(B_n \setminus C_n) = 0$ for any $n \in N_+$. Recall from above that, defining $C = \bigcup_n C_n$ and $B = \bigcup_n B_n$, $C \subset A \subset B$ and $\mu(B \setminus C) = 0$. Also note that $\{C_n\}_{n \in N_+}$ is disjoint, since $C_n \subset A_n$ for any $n \in N_+$ and the A_n are disjoint. It follows that

$$\begin{aligned}\bar{\mu}(A) = \mu(C) &= \sum_{n=1}^{\infty} \mu(C_n) && \text{(Countable Additivity)} \\ &= \sum_{n=1}^{\infty} \bar{\mu}(A_n). && \text{(Definition)}\end{aligned}$$

Therefore, $\bar{\mu}$ is a measure on $(E, \bar{\mathcal{E}})$ such that $\bar{\mu}(A) = \mu(A)$ for any $A \in \mathcal{E}$.

We can easily show that $\bar{\mathcal{E}} = \mathcal{E} \cup \mathcal{N}$. Let $A \in \mathcal{E}$. Then, $A \subset A \subset A$ and $\mu(A \setminus A) = 0$, where $A \in \mathcal{E}$, so $A \in \bar{\mathcal{E}}$. For any $N \in \mathcal{N}$, since $\emptyset \subset N \subset A$ for some $A \in \mathcal{E}$ such that $\mu(A) = 0$ (by the definition of a negligible set), where $A, \emptyset \in \mathcal{E}$ and $\mu(A \setminus \emptyset) = \mu(A) = 0$, by definition $N \in \bar{\mathcal{E}}$. Therefore, $\bar{\mathcal{E}}$ is a σ-algebra on E containing both $\mathcal{E}$ and $\mathcal{N}$; since σ-algebras are closed under finite unions, it follows that $\mathcal{E} \cup \mathcal{N} \subset \bar{\mathcal{E}}$.
Conversely, suppose $A \in \bar{\mathcal{E}}$; then, there exist $B, C \in \mathcal{E}$ such that $C \subset A \subset B$ and $\mu(B \setminus C) = 0$. Defining $N = A \setminus C$, since $A \setminus C \subset B \setminus C$, where $B \setminus C \in \mathcal{E}$ has measure 0, by definition $N \in \mathcal{N}$. Then,

$$A = C \cup (A \setminus C) = C \cup N$$

shows that $A \in \mathcal{E} \cup \mathcal{N}$, since $C \in \mathcal{E}$. This shows that $\bar{\mathcal{E}} \subset \mathcal{E} \cup \mathcal{N}$. Putting the two results together, we have

$$\bar{\mathcal{E}} = \mathcal{E} \cup \mathcal{N},$$

and the last result shows us that

$$\bar{\mu}(A) = \mu(C),$$

where $A = C \cup N$ for some $C \in \mathcal{E}$ and $N \in \mathcal{N}$.
Q.E.D.

The above theorem tells us three things:

- $\bar{\mathcal{E}} = \mathcal{E} \cup \mathcal{N}$ is itself a σ-algebra on E, and thus the smallest σ-algebra that contains both the original σ-algebra $\mathcal{E}$ and the collection of negligible sets $\mathcal{N}$.

- $\bar{\mu}$ is an extension of μ to $\bar{\mathcal{E}}$ that preserves the measure of sets in the original σ-algebra $\mathcal{E}$.

- For any $N \in \mathcal{N}$, there exists an $A \in \mathcal{E}$ such that $\mu(A) = 0$. As such, $\emptyset \subset N \subset A$, where $\emptyset \in \mathcal{E}$ and $\mu(A \setminus \emptyset) = \mu(A) = 0$, so by definition

$$\bar{\mu}(N) = \mu(\emptyset) = 0.$$

In other words, $(E, \bar{\mathcal{E}}, \bar{\mu})$ is the minimal extension of $(E, \mathcal{E}, \mu)$ to a measure space in which all negligible sets are measurable with measure 0 and the measure of the original measurable sets are preserved. We thus call $(E, \bar{\mathcal{E}}, \bar{\mu})$ the completion of $(E, \mathcal{E}, \mu)$.

3.5.3 Functions Defined Almost Everywhere

Since every measure space can be extended to its completion, we can take arbitrary measure spaces to be complete by assuming that we are working with their completion. For this reason, going forward we will assume that $(E, \mathcal{E}, \mu)$ is complete.
Let us return again to the almost everyhwere equivalence problem. For any $\mathcal{E}$-measurable real function f and a function $g : E \to \mathbb{R}$ that is μ-a.e. equivalent to f, there exists an $A \in \mathcal{E}$ such that $f = g$ on A and $\mu(A^c) = 0$. Since $(E, \mathcal{E}, \mu)$ is now complete, g is also $\mathcal{E}$-measurable; to see this, choose any $B \in \mathcal{B}(\mathbb{R})$ and observe that

$$\begin{aligned} g^{-1}(B) &= \left(g^{-1}(B) \cap A\right) \cup \left(g^{-1}(B) \cap A^c\right) \\ &= \left(f^{-1}(B) \cap A\right) \cup \left(g^{-1}(B) \cap A^c\right), \end{aligned}$$

since $f = g$ on A. By the measurability of f, $f^{-1}(B) \cap A \in \mathcal{E}$, while $g^{-1}(B) \cap A^c$, being a subset of a measure 0 set A^c, is a negligible set and thus $\mathcal{E}$-measurable. It follows that $g^{-1}(B) \in \mathcal{E}$, so that g is a $\mathcal{E}$-measurable real valued function.

Now we are able to extend a function defined almost everywhere on E any way we want, and still retain measurability. Let $A \in \mathcal{E}$ be a set such that $\mu(A^c) = 0$, and $f : A \to \mathbb{C}$ a $\mathcal{E}$-measurable function. By definition, its trivial extension $\bar{f}$ is a $\mathcal{E}$-measurable complex function. Let $g : E \to \mathbb{C}$ be an arbitrary extension of f to $\mathbb{E}$, so that $g(x) = f(x)$ for any $x \in A$. This means that $\bar{f} = g$ a.e. $[\mu]$, which in turn implies that the real and imaginary parts of $\bar{f}$ and g are almost everywhere equivalent. By the measurability of $\bar{f}$, the real and imaginary parts of g are also $\mathcal{E}$-measurable, which means that g is a $\mathcal{E}$-measurable complex function.
In addition, because of the equivalence of the integral of almost everywhere equivalent functions, if f is μ-integrable ($=\bar{f}$ is μ-integrable), then so is g, and

$$\int_E f d\mu = \int_E \bar{f} d\mu = \int_E g d\mu.$$

This implies that the value of the integral of f does not depend on the specific way in which f is extended to E, and therefore that we do not need to specify the extension of f at all when talking about the integral or measurability of f.

The following are versions of the MCT, MCT for series, and DCT for functions defined almost everywhere, which are proved through extensive use of the trivial extensions of functions defined almost everywhere. We also state a DCT analogue for the MCT for series:

Theorem 3.12 (Almost Everywhere Version of the MCT)
Let $(E,\mathcal{E},\mu)$ be a complete measure space, and $\{f_n\}_{n\in N_+}$ a sequence in $\mathcal{E}_+$ defined almost everywhere on E. Suppose that

$$f_n(x) \le f_{n+1}(x)$$

for almost every $x \in E$. Then, the limit

$$f_n(x) \nearrow f(x)$$

exists for almost every $x \in E$, f is a measurable function defined almost everywhere on E, and

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu.$$

Proof) For any $n \in N_+$, let

$$f_n(x) \le f_{n+1}(x)$$

for any $x \in A_n$ such that $\mu(A_n^c) = 0$. Then, defining $A = \bigcap_n A_n$,

$$\mu(A^c) = \mu\left(\bigcup_n A_n^c\right) \le \sum_{n=1}^{\infty} \mu(A_n^c) = 0$$

by countable subadditivity, so that A^c is a negligible set. We can thus treat each f_n as a function defined on A.

For any $x \in A$, since $\{f_n(x)\}_{n\in N_+}$ is an increasing sequence of non-negative real numbers, it has a limit $f_x \in [0,+\infty]$. We can then define the function $f : A \to [-\infty,+\infty]$ as $f(x) = f_x$ for any $x \in A$. Let $\bar{f}_n$ be the trivial extension of f_n to E for any $n \in N_+$ and $\bar{f}$ the trivial extension of f to E. Then, since

$$\lim_{n\to\infty} \bar{f}_n(x) = \begin{cases} f(x) & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} = \bar{f}(x)$$

for any $x \in E$ and $\{\bar{f}_n\}_{n\in N_+}$ is a sequence of increasing $\mathcal{E}$-measurable functions, $\bar{f}$ is a $\mathcal{E}$-measurable non-negative function because measurability is preserved across limits; by definition, this also implies that f is $\mathcal{E}$-measurable function.

By the MCT, we can now see that

$$\begin{aligned}
\int_E f d\mu &= \int_E \bar{f} d\mu \\
&= \lim_{n\to\infty} \int_E \bar{f}_n d\mu && \text{(The MCT)} \\
&= \lim_{n\to\infty} \int_E f_n d\mu,
\end{aligned}$$

where the first and third equalities are justified by the definition of a measurable function defined almost everywhere.
Q.E.D.

Theorem 3.13 (Almost Everywhere Version of the MCT for Series)
Let $(E, \mathcal{E}, \mu)$ be a complete measure space, and $\{f_n\}_{n\in N_+}$ a sequence in $\mathcal{E}_+$ only defined almost everywhere on E. Then,

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is defined for almost every $x \in E$, f is a measurable function, and

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Proof) For any $n \in N_+$, let f_n be defined on a set $A_n \in \mathcal{E}$ such that $\mu(A_n^c) = 0$. Then, defining $A = \bigcap_n A_n$,

$$\mu(A^c) = \mu\left(\bigcup_n A_n^c\right) \le \sum_{n=1}^{\infty} \mu(A_n^c) = 0$$

by countable subadditivity, so that A^c is a negligible set. We can thus treat each f_n as a function defined only on A.

Letting $\bar{f}_n$ be the trivial extensions of f_n to E, the sequence $\{\bar{f}_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable non-negative functions, and by the MCT for series,

$$\bar{f} = \sum_{n=1}^{\infty} \bar{f}_n$$

is a measurable non-negative function and

$$\int_E \bar{f} d\mu = \sum_{n=1}^{\infty} \int_E \bar{f}_n d\mu.$$

Defining $f : A \to [0, +\infty]$ as

$$f(x) = \bar{f}(x) = \sum_{n=1}^{\infty} \bar{f}_n(x) = \sum_{n=1}^{\infty} f_n(x)$$

for any $x \in A$, $\bar{f}$ is the trivial extension of f to E, and because $\bar{f}$ is measurable, so is f. In addition, by the definition of the integral of functions defined almost everywhere,

$$\int_E f d\mu = \int_E \bar{f} d\mu = \sum_{n=1}^{\infty} \int_E \bar{f}_n d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Q.E.D.

Theorem 3.14 (Almost Everywhere Version of the DCT)
Let $(E, \mathcal{E}, \mu)$ be a complete measure space, and $\{f_n\}_{n \in N_+}$ a sequence of $\mathcal{E}$-measurable complex functions only defined almost everywhere on E. Suppose that

- There exists a function $g \in \mathcal{E}_+$ defined almost everywhere on E such that $\int_E g d\mu < +\infty$ and, for any $n \in N_+$, $|f_n(x)| \leq g(x)$ for almost every $x \in E$.
- The limit

  $$\lim_{n \to \infty} f_n(x) = f(x)$$

  exists for almost every $x \in E$.

Then, $f \in L^1(\mathcal{E}, \mu)$ and

$$\int_E f d\mu = \lim_{n \to \infty} \int_E f_n d\mu.$$

Proof) For any $n \in N_+$, let

$$|f_n(x)| \leq g(x)$$

for any $x \in A_n$, where $\mu(A_n^c) = 0$, and suppose

$$\lim_{n \to \infty} f_n(x) = f(x)$$

for any $x \in B$, where $\mu(B^c) = 0$. Then, defining $A = (\bigcap_n A_n) \cap B$,

$$\mu(A^c) = \mu\left(\left(\bigcup_n A_n^c\right) \cup B^c\right) \leq \sum_{n=1}^{\infty} \mu(A_n^c) + \mu(B^c) = 0$$

by countable subadditivity, so that A^c is a negligible set. We can thus treat each f_n, f and g as a function defined only on A.

Letting $\bar{f}_n$, $\bar{f}$ and $\bar{g}$ be the trivial extensions of their respective functions defined on A, note that $\{\bar{f}_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable complex functions, g a non-negative function in $\mathcal{E}_+$, and

$$\lim_{n\to\infty} \bar{f}_n(x) = \bar{f}(x)$$

for any $x \in E$. By the preservation of measurability across limits, $\bar{f}$ is also a $\mathcal{E}$-measurable complex function, indicating that the function f defined only on A is also measurable. Furthermore,

$$\left|\bar{f}_n(x)\right| = \begin{cases} |f_n(x)| & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \leq \bar{g}(x)$$

for any $x \in E$, and $\int_E \bar{g} d\mu = \int_E g d\mu < +\infty$.
By the DCT, we can see that $\bar{f} \in L^1(\mathcal{E},\mu)$, which implies $f \in L^1(\mathcal{E},\mu)$, and

$$\int_E f d\mu = \int_E \bar{f} d\mu = \lim_{n\to\infty} \int_E \bar{f}_n d\mu = \lim_{n\to\infty} \int_E f d\mu.$$

Q.E.D.

Theorem 3.15 (DCT for Series)
Let $(E,\mathcal{E},\mu)$ be a complete measure space, and $\{f_n\}_{n\in N_+}$ a sequence of $\mathcal{E}$-measurable complex functions defined almost everywhere on E such that

$$\sum_{n=1}^{\infty} \int_E |f_n| d\mu < +\infty.$$

Then, the series $\sum_{n=1}^{\infty} f_n(x)$ converges for almost every $x \in E$, and defining f as

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

for the $x \in E$ such that the right hand side converges, $f \in L^1(\mathcal{E},\mu)$ and

$$\int_E f d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu.$$

Proof) The sequence $\{|f_n|\}_{n\in N_+}$ is a sequence of measurable non-negative functions defined almost everywhere on E. By the MCT for series, the limit

$$h(x) = \sum_{n=1}^{\infty} |f_n(x)|$$

is defined for almost every $x \in E$, h is a measurable non-negative valued function, and

$$\int_E h d\mu = \sum_{n=1}^{\infty} \int_E |f_n| d\mu < +\infty,$$

where the last inequality follows by assumption. Letting $\bar{h}$ be the trivial extension of h to E, this implies that

$$\int_E \bar{h} d\mu = \int_E h d\mu < +\infty,$$

so that, by the vanishing property of the integration of non-negative functions, $B = \{x \in E \mid \bar{h}(x) < +\infty\} \in \mathcal{E}$ satisfies $\mu(B^c) = 0$. Letting $H \in \mathcal{E}$ be the set on which h is defined, where $\mu(H^c) = 0$,

$$h(x) = \bar{h}(x) = \sum_{n=1}^{\infty} |f_n(x)| < +\infty$$

for any $x \in H \cap B$, that is, the series $\sum_{n=1}^{\infty} f_n(x)$ converges to a value in $\mathbb{C}$ for any $x \in H \cap B$. $\mu((H \cap B)^c) = 0$ by subadditivity, so $\sum_{n=1}^{\infty} f_n(x)$ converges for almost every $x \in E$.

For any $n \in N_+$, let f_n be defined on a set $A_n \in \mathcal{E}$ such that $\mu(A_n^c) = 0$. Then, defining $A = (\bigcap_n A_n) \cap B \cap H$,

$$\mu(A^c) = \mu\left(\left(\bigcup_n A_n^c\right) \cup B^c \cup H^c\right) \leq \sum_{n=1}^{\infty} \mu(A_n^c) + \mu(B^c) + \mu(H^c) = 0$$

by countable subadditivity, so that A^c is a negligible set. We can thus treat each f_n and h as a function defined only on A.

From the result we showed above, we can define $f : A \to \mathbb{C}$

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

for any $x \in A$. Let $\bar{f}$, $\bar{f}_n$, $\bar{h}$ be the trivial extensions of the associated functions to E. Then,

$$\bar{f}(x) == \begin{cases} \sum_{n=1}^{\infty} f_n(x) & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} = \sum_{n=1}^{\infty} \bar{f}_n(x)$$

for any $x \in E$.

Defining the sequence $\{g_n\}_{n\in N_+}$ of functions on E as

$$g_n(x) = \sum_{i=1}^{n} \bar{f}_i(x)$$

for any $x \in E$, by the measurability of each $\bar{f}_i$ and the preservation of measurability across sums, the function g_n is a measurable complex function. Furthermore,

$$|g_n(x)| \le \sum_{i=1}^{n} \left|\bar{f}_i(x)\right| \le \sum_{i=1}^{\infty} \left|\bar{f}_i(x)\right| = \begin{cases} \sum_{i=1}^{\infty} |f_i(x)| & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} = \bar{h}(x)$$

for any $x \in E$, where $\bar{h}$ is a non-negative measurable function such that

$$\int_E \bar{h} d\mu = \int_E h d\mu < +\infty.$$

Finally,

$$\bar{f}(x) = \sum_{n=1}^{\infty} \bar{f}_n(x) = \lim_{n\to\infty} \sum_{i=1}^{n} \bar{f}_i(x) = \lim_{n\to\infty} g_n(x)$$

for any $x \in E$. Since measurability is preserved across limits, $\bar{f}$ is measurable, which means that f is also measurable.
Therefore, by the DCT, $\bar{f} \in L^1(\mathcal{E}, \mu)$, which implies $f \in L^1(\mathcal{E}, \mu)$, and

$$\begin{aligned} \int_E f d\mu = \int_E \bar{f} d\mu &= \lim_{n\to\infty} \int_E g_n d\mu \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \int_E \bar{f}_i d\mu \\ &= \sum_{n=1}^{\infty} \int_E \bar{f}_n d\mu \\ &= \sum_{n=1}^{\infty} \int_E f_n d\mu. \end{aligned} \tag{DCT}$$

Q.E.D.

3.6 Transition Kernels

In this section we introduce and study transition kernels, which are mathematical objects that form the basis of the concept of conditional probabilities in probability theory. First, we state a result concerning the characterization of integrals that will prove useful going forward.

3.6.1 An Elementary Representation Theorem

So far, we have seen that the integral with respect to some measure is a linear functional on the vector space of all complex functions (this point will be studied further in the next chapter on Borel measures) and that the MCT holds for integrals of non-negative functions. The following theorem shows that these properties also fully characterize integration, in the sense that, for any linear functional on the vector space of all complex functions that also satisfies the MCT, there exists a unique measure such that the integral of the function with respect to that measure is precisely the value of the linear functional for that function.
This can be seen as a stronger version of the Riesz representation theorem, which will be introduced in the next chapter, where the condition that the underlying space be a locally compact Hausdorff space is replaced with the requirement that the linear functional in question satisfies the MCT. The formal statement is given as follows:

Theorem 3.16 (Characterization of Integrals)
Let $(E, \mathcal{E})$ be a measurable space, and $\Lambda : \mathcal{E}_+ \to [0, +\infty]$ a function such that

i) $\Lambda(I_\emptyset) = 0$.

ii) For any $a \in [0, +\infty)$ and $f, g \in \mathcal{E}_+$,

$$\Lambda(af + g) = a \cdot \Lambda f + \Lambda g.$$

iii) For any increasing sequence of non-negative measurable functions $\{f_n\}_{n \in N_+}$,

$$\Lambda f_n \nearrow \Lambda f,$$

as $n \to \infty$, where f is the pointwise limit of $\{f_n\}_{n \in N_+}$.

Then, there exists a unique measure μ on $(E, \mathcal{E})$ such that

$$\Lambda f = \int_E f d\mu$$

for any $f \in \mathcal{E}_+$.

Proof) For any $A \in \mathcal{E}$, we will define the function $\mu : \mathcal{E} \to [0, +\infty]$ as

$$\mu(A) = \Lambda(I_A).$$

We will now show that μ is a measure on $(E,\mathcal{E})$, and that the integral of any non-negative measurable function with respect to μ equals its value under the linear functional Λ.

It is immediately clear that

$$\mu(\emptyset) = \Lambda(I_\emptyset) = 0.$$

Furthermore, for any disjoint sequence of measurable sets $\{A_n\}_{n\in N_+}$, letting $A = \bigcup_n A_n$, we have $I_A = \sum_{n=1}^{\infty} I_{A_n}$; since $\{\sum_{i=1}^{n} I_{A_i}\}_{n\in N_+}$ is an increasing sequence of non-negative measurable functions with limit I_A, by the linearity and MCT properties of Λ, we have

$$\begin{aligned}\mu(A) = \Lambda(I_A) = \lim_{n\to\infty} \Lambda\left(\sum_{i=1}^{n} I_{A_i}\right) &= \lim_{n\to\infty} \sum_{i=1}^{n} \Lambda(I_{A_n}) \\ &= \sum_{n=1}^{\infty} \Lambda(I_{A_n}) = \sum_{n=1}^{\infty} \mu(A_n),\end{aligned}$$

where the series on the right hand side converges in $[0,+\infty]$ because each $\mu(A_n)$ is non-negative. Therefore, μ is countably additive and a measure on $(E,\mathcal{E})$.

Now let f be a measurable simple function with canonical form

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$$

for some $\alpha_1,\cdots,\alpha_n \in [0,+\infty)$ and $A_1,\cdots,A_n \in \mathcal{E}$. Then,

$$\begin{aligned}\int_E f d\mu = \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i) &= \sum_{i=1}^{n} \alpha_i \cdot \Lambda(I_{A_i}) \\ &= \Lambda\left(\sum_{i=1}^{n} \alpha_i \cdot I_{A_i}\right) && \text{(By the Linearity of } \Lambda) \\ &= \Lambda f.\end{aligned}$$

Now let f be a general non-negative measurable function. Then, letting $\{f_n\}_{n\in N_+}$ be a sequence of measurable simple functions increasing to f, by the MCT and the MCT property of Λ we have

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu = \lim_{n\to\infty} \Lambda f_n = \Lambda f.$$

Therefore, for any $f \in \mathcal{E}_+$, we have

$$\Lambda f = \int_E f d\mu \in [0,+\infty].$$

The uniqueness of μ follows easily. Suppose that v is another measure on $(E,\mathcal{E})$ such that

$$\Lambda f = \int_E f dv$$

for any $f \in \mathcal{E}_+$. Then, for any $A \in \mathcal{E}$, since $I_A \in \mathcal{E}_+$, we have

$$\mu(A) = \Lambda(I_A) = v(A),$$

so that $\mu = v$ on $\mathcal{E}$.
Q.E.D.

The measure μ represents the linear functional Λ in the sense that the value of a function under Λ is its integral with respect to μ. This same sense of representation will be used in the next chapter for the Riesz representation theorem, except there the proof will be complicated due to the relevant linear functional being defined only on the space of continuous functions.

3.6.2 Definitions and Measure-Kernel-Functions

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A transition kernel K from $(E, \mathcal{E})$ into $(F, \mathcal{F})$ is a function $K : E \times \mathcal{F} \to [0, +\infty]$ such that:

- For any $A \in \mathcal{F}$, $K(\cdot, A)$ is a $\mathcal{E}$-measurable non-negative function.
- For any $x \in E$, $K(x, \cdot)$ is a measure on $(F, \mathcal{F})$.

K is called a transition probability kernel if $K(x, \cdot)$ is a probability measure for each $x \in E$, and a Markov kernel on $(E, \mathcal{E})$ if it is a transition probability kernel from $(E, \mathcal{E})$ into itself. We denote the integral of a function $f \in \mathcal{F}_+$ with respect to the measure $K(x, \cdot)$ for some $x \in E$ by

$$\int_F f(y) K(x, dy).$$

The usefulness of transition kernels will be made immediately clear if we let K e a Markov kernel on $(E, \mathcal{E})$, E a finite space and $\mathcal{E}$ the corresponding discrete σ-algebra. Letting $E = \{x_1, \cdots, x_n\}$, in this case we can construct an $m \times n$ matrix P with (i, j)th element equal to $K(x_i, \{x_j\})$. Then, since $K(x_i, \cdot)$ is a probability measure for each $1 \leq i \leq n$, the matrix P can be regarded as a transition probability matrix of some finite Markov process, and the value $K(x_i, \{x_j\})$ specifically the probability of the process taking the value x_j given that the previous iteration took the value x_i. Therefore, Markov kernels on arbitrary measurable spaces can be seen as generalizations of the concept of the transition probability matrix of finite Markov processes, and transition kernesl the generalization of Markov kernels to arbitrary measures.

The following theorem shows that we can construct measurable functions and measures using transition kernels:

Theorem 3.17 Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces, and K a transition kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$. Then, the following hold true:

i) The transformation $T_K : \mathcal{F}_+ \to \mathcal{E}_+$ defined as

$$(T_K f)(x) = \int_F f(y)K(x,dy)$$

for any $x \in E$ is

- **Linear**:
  For any $a \in [0,+\infty)$ and $f,g \in \mathcal{F}_+$,

$$T_K(af+g) = a \cdot T_K f + T_K g$$

- **Continuous under Increasing Limits**:
  For any increasing sequence $\{f_n\}_{n\in N_+} \subset \mathcal{F}_+$ with pointwise limit f,

$$T_K f_n \nearrow T_K f$$

  as $n \to \infty$.

ii) For any measure μ on $(E,\mathcal{E})$, the function $\mu K : \mathcal{F} \to [0,+\infty]$ defined as

$$(\mu K)(A) = \int_E K(x,A)d\mu(x)$$

for any $A \in \mathcal{F}$ is a measure on $(F,\mathcal{F})$.

iii) For any measure μ on $(E,\mathcal{E})$ and non-negative measurable function $f \in \mathcal{F}_+$,

$$\int_F f d(\mu K) = \int_E (T_K f)d\mu,$$

and we denote the above quantity by $\int_E \int_F f(y)K(x,dy)d\mu(x)$.

Proof) We will first show that the operator T_K is linear and continuous under increasing limits. For any $a \in [0,+\infty)$ and $f,g \in \mathcal{F}_+$, for any $x \in E$,

$$\begin{aligned}(T_K(af+g))(x) &= \int_F (a \cdot f(y) + g(y))K(x,dy)\\ &= a \cdot \int_F f(y)K(x,dy) + \int_F g(y)K(x,dy) = a \cdot (T_K f)(x) + (T_K g)(x)\end{aligned}$$

by the linearity of integration of non-negative functions, so it holds that

$$T_K(af+g) = T_K f + T_K g$$

on E. Furthermore, for any increasing sequence of functions $\{f_n\}_{n\in N_+} \subset \mathcal{F}_+$ with limit f, for any $x \in E$

$$(T_K f)(x) = \int_F f(y)K(x,dy) = \lim_{n\to\infty} \int_F f_n(y)K(x,dy) = \lim_{n\to\infty} (T_K f_n)(x)$$

by the MCT, and by the monotonicity of integration, for any $n \in N_+$,

$$(T_K f_n)(x) = \int_F f_n(y)K(x,dy) \le \int_F f_{n+1}(y)K(x,dy) = (T_K f_{n+1})(x).$$

Therefore, $\{T_K f_n\}_{n\in N_+}$ is a sequence of non-negative functions on E increasing pointwise to $T_K f$.

It remains to show that $T_K f$ is measurable for any $f \in \mathcal{F}_+$. Let f be a measurable simple function with canonical form $f = \sum_{i=1}^n \alpha_i \cdot I_{A_i}$ for $\alpha_1, \cdots, \alpha_n \in [0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{F}$. Then, by linearity,

$$T_K f = \sum_{i=1}^n \alpha_i \cdot T_K I_{A_i} = \sum_{i=1}^n \alpha_i \cdot K(\cdot, A_i).$$

Since $K(\cdot, A_i) \in \mathcal{E}_+$ for each $1 \le i \le n$ by definition, it follows that $T_K f \in \mathcal{E}_+$ as well.

Now let $f \in \mathcal{F}_+$ in general. Then, letting $\{f_n\}_{n\in N_+}$ be a sequence of measurable simple functions increasing to f, by the continuity of T_K under increasing limits we have

$$T_K f = \lim_{n\to\infty} T_K f_n,$$

and because $T_K f_n \in \mathcal{E}_+$ for any $n \in N_+$, $T_K f$ is also $\mathcal{E}$-measurable. Therefore, T_K is a linear transformation mapping from $\mathcal{F}_+$ into $\mathcal{E}_+$ that is continuous under increasing limits.

Now let μ be a measure on $(E, \mathcal{E})$, and define the mapping $\Lambda : \mathcal{F}_+ \to [0, +\infty]$ as

$$\Lambda f = \int_E (T_K f) d\mu$$

for any $f \in \mathcal{F}_+$, where the integral on the right hand side is well-defined because $T_K f \in \mathcal{E}_+$. We will show that Λ is a linear functional with the MCT property:

- $\Lambda(I_\emptyset) = \int_E (T_K I_\emptyset) d\mu = \int_E K(x, \emptyset) d\mu(x) = 0$ because $K(x, \emptyset) = 0$ for any $x \in E$.

– For any $a \in [0, +\infty)$ and $f, g \in \mathcal{F}_+$,

$$\Lambda(af+g) = \int_E T_K(af+g)d\mu = \int_E (a \cdot T_K f + T_K g)\, d\mu \qquad \text{(Linearity of } T\text{)}$$
$$= a \cdot \int_E (T_K f) d\mu + \int_E (T_K g) d\mu = a \cdot \Lambda f + \Lambda g.$$
$$\text{(Linearity of Integration)}$$

– For any increasing sequence $\{f_n\}_{n \in N_+} \subset \mathcal{F}_+$ with pointwise limit f, because $T_K f_n \nearrow T_K f$ as $n \to \infty$ by the continuity of T_K under increasing limits, $\{T_K f_n\}_{n \in N_+}$ is an increasing sequence of $\mathcal{E}$-measurable non-negative functions with limit $T_K f$: by the MCT,

$$\Lambda f = \int_E (T_K f) d\mu = \lim_{n \to \infty} \int_E (T_K f_n)\, d\mu = \lim_{n \to \infty} \Lambda f_n,$$

and by the monotonicity of integration, $\Lambda f_n \leq \Lambda f_{n+1}$ for any $n \in N_+$.

Therefore, by theorem 3.16, there exists a unique measure v on $(F, \mathcal{F})$ such that

$$\Lambda f = \int_F f dv$$

for any $f \in \mathcal{F}_+$. For any $A \in \mathcal{F}$,

$$v(A) = \int_F (I_A) dv = \Lambda(I_A) = \int_E (T_K I_A) d\mu = \int_E K(x, A) d\mu(x) = (\mu K)(A).$$

Therefore, $v = \mu K$ on $\mathcal{F}$ and μK is a measure on $(F, \mathcal{F})$. In addition, for any $f \in \mathcal{F}_+$, we have

$$\int_F f d(\mu K) = \Lambda f = \int_E (T_K f) d\mu.$$

Q.E.D.

The quantity

$$\int_E \int_F f(y) K(x, dy) d\mu(x)$$

for any measure μ on $(E, \mathcal{E})$ and function $f \in \mathcal{F}_+$ allows us to, in a sense, "integrate" a function on F with respect to a measure on $(E, \mathcal{E})$. This will prove especially useful in our study of Markov chains.

3.6.3 Propreties of Markov Kernels

Let $(E,\mathcal{E})$, $(F,\mathcal{F})$ and $(G,\mathcal{G})$ be measurable spaces, and K,L transition kernels from $(E,\mathcal{E})$ into $(F,\mathcal{F})$ and $(F,\mathcal{F})$ into $(G,\mathcal{G})$, respectively. Much like functions, we can think of compositions of transition kernels. Formally, the product of K and L is a function $KL: E \times \mathcal{G} \to [0,+\infty]$ defined as

$$(KL)(x,A) = \int_F L(y,A)K(x,dy)$$

for any $(x,A) \in E \times \mathcal{G}$. Because

- For any $A \in \mathcal{G}$, $L(\cdot,A) \in \mathcal{F}_+$ and $(KL)(x,\cdot) = T_K L(\cdot,A) \in \mathcal{E}_+$ by the above theorem, and
- For any $x \in E$, $\mu = K(x,\cdot)$ is a measure on $(F,\mathcal{F})$, so that $(KL)(x,\cdot) = \mu L$ is a measure on $(G,\mathcal{G})$ by the above theorem,

by definition KL is a transition kernel from $(E,\mathcal{E})$ into $(G,\mathcal{G})$. Note that, if both K and L are transition probability kernels, then for any $x \in E$

$$(KL)(x,G) = \int_F L(y,G)K(x,dy) = K(x,F) = 1$$

and thus KL is itself a transition probability kernel.

The concept of the product of transition kernels has especially important implications for Markov kernels, since it allows us to define the power of Markov kernels.
Let $(E,\mathcal{E})$ be a measurable space and P a Markov kernel on $(E,\mathcal{E})$. Heuristically, we may think of P as a transition probability for some Markov chain. In accordance with the above definition of the product of transition kernels, we can define P^n recursively as follows:

- $P^1 = P$, and
- $P^n = P^{n-1}P$ for any $n \geq 2$.

The reason for defining P^n as above instead of the other way around (PP^{n-1}) will be made clearer in the next section. Note that, for any $n \geq 2$, by definition,

$$P^n(x,A) = \int_E P(y,A)P^{n-1}(x,dy)$$

for any $(x,A) \in E \times \mathcal{E}$.

Defining the kernel I as

$$I(x,A) = \delta_x(A),$$

where δ_x is the Dirac delta measure on $(E,\mathcal{E})$ sitting at x, for any $(x,A) \in E \times \mathcal{E}$, I is clearly a Markov kernel on $(E,\mathcal{E})$, and we can see that

$$(P \cdot I)(x,A) = \int_E I(y,A)P(x,dy) = \int_E I_A(y)P(x,dy) = P(x,A),$$

so that the product of P and I is P. It is also easy to see that $I \cdot P = P$;

$$(I \cdot P)(x,A) = \int_E P(y,A)I(x,dy) = \int_E P(y,A)d\delta_x(y) = P(x,A).$$

Therefore, we denote call I the identity kernel and denote $I = P^0$. This has the advantage that we can express

$$P^n(x,A) = \int_E P(y,A)P^{n-1}(x,dy)$$

for any $n \in N_+$, and not just $n \geq 2$.

By the theorem proved in the previous section, we can obtain a convenient expression for integrals with respect to the product P^n. For any $f \in \mathcal{E}_+$, $n \in N_+$ and $x \in E$, denote $P^{n-1}(x,\cdot) = \mu_x$. Then, because $P^n(x,\cdot)$ satisfies

$$P^n(x,A) = \int_E P(y,A)P^{n-1}(x,dy) = \int_E P(y,A)d\mu_x(y),$$

in the notation of the above theorem, $P^n(x,\cdot) = \mu_x P$, and it follows that

$$\begin{aligned}\int_E f(y)P^n(x,dy) = \int_E fd(\mu_x P) &= \int_E \int_E f(z)P(y,dz)d\mu_x(y) \\ &= \int_E \int_E f(z)P(y,dz)P^{n-1}(x,dy) = \int_E (T_P f)(y)P^{n-1}(x,dy).\end{aligned}$$

Therefore, continuing to rewrite the integral in this way shows us that the integral of f with respect to P^n can be evaluated sequentially through integrals with respect to P.

Some Markov kernels have special properties that make them especially attractive in probabilistic analysis. Here we state some of those properties, which will be heuristically expanded upon in the next section. Throughout, we let $(E,\mathcal{E})$ be a measurable space and P a Markov kernel on $(E,\mathcal{E})$.

- **Stationary Distributions**

  Let π be a measure on $(E,\mathcal{E})$. Then, π is a stationary distribution of P if

  $$\int_E P(x,A)d\pi(x) = \pi(A)$$

  for any $A \in \mathcal{E}$. Note that, in the notation of theorem 3.17, this means that $\pi P = \pi$.
  π is called "stationary" for P because, for any $n \geq 2$, assuming that $\pi P^{n-1} = \pi$ for any $A \in \mathcal{E}$, we have

  $$\begin{aligned}\int_E P^n(x,A)d\pi(x) &= \int_E (T_{P^{n-1}}P(\cdot,A))(x)d\pi(x) && \text{(By definition)}\\ &= \int_E P(x,A)d(\pi P^{n-1})(x) && \text{(Theorem 3.17)}\\ &= \int_E P(x,A)d\pi(x) && \text{(Inductive Hypothesis)}\\ &= \pi(A) && \text{(Definition of Stationarity)}\end{aligned}$$

  for any $A \in \mathcal{E}$, so that $\pi P^n = \pi$ on $\mathcal{E}$. By induction, $\pi P^n = \pi$, that is,

  $$\int_E P^n(x,A)d\pi(x) = \pi(A)$$

  for any $A \in \mathcal{E}$, for any $n \in N_+$. Heuristically, this means that the Markov chain with P has unconditional distribution equal to π throughout the chain, regardless of the initial value x.

- **Reversibility**

  We say that P is reversible with respect to a measure π on $(E,\mathcal{E})$ if, for any $A,B \in \mathcal{E}$,

  $$\int_A P(y,B)d\pi(x) = \int_B P(y,A)d\pi(y).$$

  Heuristically, reversibility implies that the order of the Markov chain with P as its transition probability can be reversed but still maintain the same probabilistic structure.
  If P is reversible with respect to π, then π is stationary for P. To see this, choose any $A \in \mathcal{E}$; then,

  $$\begin{aligned}\int_E P(x,A)d\pi(x) &= \int_A P(x,E)d\pi(x) && \text{(Reversibility)}\\ &= \pi(A), && \text{(P is a Markov Kernel)}\end{aligned}$$

  and by definition π is stationary for P.
  The most important application of the above result is that the Markov kernel used in the

Metropolis-Hastings algorithm is reversible with respect to the target distribution, which is why the algorithm converges to that target distribution. This will be shown in greater detail in the next section.

- **ϕ-Irreducibility**
  Let ϕ be a measure on $(E, \mathcal{E})$. Then, we say that P is ϕ-irreducible if, for any $x \in E$ and $A \in \mathcal{E}$ such that $\phi(A) > 0$, there exists an $n \in N_+$ such that

  $$P^n(x, A) > 0.$$

  Heuristically, it means that the Markov chain with P as its transition probability eventually enters the set A with positive probability regardless of the initial value x. As such, this means that the chain will pass through every set to which ϕ assigns a positive probability. Usually, this ϕ is taken to be the stationary distribution of P.

- **Aperiodicity**
  Let π be the stationary distribution of P. We say that P is periodic if there exists an integer $d \geq 2$ and disjoint sets $A_1, \cdots, A_d \in \mathcal{E}$ such that $P(x, A_{i+1}) = 1$ for any $x \in A_i$ and $1 \leq i \leq d-1$, $P(x, A_1) = 1$ for any $x \in A_d$, and $\pi(A_1) > 0$. Suppose that $\pi(A_i) > 0$ for some $1 \leq i \leq d-1$; then,

  $$\pi(A_{i+1}) = \int_E P(x, A_{i+1}) d\pi(x) = \int_{A_i} P(x, A_{i+1}) d\pi(x) = \pi(A_i) > 0,$$

  so that, by induction, $\pi(A_1) > 0$ implies $\pi(A_d) = \cdots = \pi(A_1) > 0$.
  Heuristically, periodicity indicates that the Markov chain with P as its transition probability periodically circulates through a collection of disjoint sets; this means that the chain is predictable to a degree.
  P is said to be aperiodic if P is not periodic.

Chapter 4

Borel Measures

In this chapter we focus on function spaces as vector spaces. Specifically, letting $(E,\mathcal{E})$ be a measurable space and $\mathcal{F}(E,\mathcal{E})$ the set of all complex $\mathcal{E}$-measurable functions, then $\mathcal{F}(E,\mathcal{E})$ is a vector space over the complex field. To see this, recall that the set of all complex functions defined on E forms a vector space over the complex field. Therefore, we need only establish that $\mathcal{F}(E,\mathcal{E})$ is closed under the pointwise addition and scalar multiplication operations, and that it contains the zero function. However, these follow easily from the facts that constant functions are measurable and that measurability is preserved across addition and scalar multiplication.

Since the pair $\mathcal{F}(E,\mathcal{E})$ is a vector space over the complex field, we can now define linear functionals $\Lambda : \mathcal{F}(E,\mathcal{E}) \to \mathbb{C}$, that is, linear transformations from $\mathcal{F}(E,\mathcal{E})$ into the complex field. There is a very close relationship between linear functionals and integration.
Specifically, consider any measure μ on $(E,\mathcal{E})$. $L^1(\mathcal{E},\mu)$ is a linear subspace of $\mathcal{F}(E,\mathcal{E})$, since it contains the zero function, and if $f,g \in \mathcal{F}(E,\mathcal{E})$ are μ-integrable, then for any $a \in \mathbb{C}$, $af+g \in \mathcal{F}(E,\mathcal{E})$ and

$$\int_E |af+g|d\mu \leq |a| \cdot \int_E |f|d\mu + \int_E |g|d\mu < +\infty$$

by the monotonicity and linearity of integration, which tells us that $af+g \in L^1(\mathcal{E},\mu)$.
Now define the function $\Lambda : L^1(\mathcal{E},\mu) \to \mathbb{C}$ as

$$\Lambda(f) = \int_E f d\mu.$$

for any $f \in L^1(\mathcal{E},\mu)$. Then, for any $a \in \mathbb{C}$ and $f,g \in L^1(\mathcal{E},\mu)$, by the linearity of integration

$$\Lambda(af+g) = \int_E (af+g)d\mu = a \cdot \int_E f d\mu + \int_E g d\mu = a \cdot \Lambda(f) + \Lambda(g),$$

which tells us that Λ is a linear transformation and thus a linear functional. Note also that this linear functional is positive, in that $\Lambda(f) \geq 0$ if $f \geq 0$ (the integral of non-negative functions is always non-negative).

The focus of this chapter is the Riesz Representation Theorem, which tells us that the converse of the above claim holds. Formally, letting $(E,\mathcal{E})$ be a Borel space with certain topological properties, for any positive linear functional Λ defined on the subset of all continuous complex functions on E, there exists a measure μ on $(E,\mathcal{E})$ such that

$$\Lambda(f) = \int_E f d\mu$$

for any continuous complex function f on E. This powerful result also implies the existence of the Lebesgue measure on euclidean spaces.
This theorem makes extensive use of the topological properties of the Borel space $(E,\mathcal{E})$ laid out in chapter 1, especially Urysohn's lemma.

4.1 The Riesz Representation Theorem

We now state the theorem first, before making a few remarks and then moving onto the proof. The theorem is formally called the Riesz-Markov-Kakutani representation, since Riesz first stated it for continuous functions on the unit interval $[0,1]$, Markov extended it to continuous functions on non-compact spaces and Kakutani finally generalized the result to hold on locally compact Hausdorff spaces.

Theorem 4.1 (Riesz-Markov-Kakutani Representation Theorem)
Let (E,τ) be a locally compact Hausdorff space and $\mathcal{B}(E,\tau)$ the corresponding Borel σ-algebra. For any positive linear functional Λ on $C_c(E,\tau)$, there exists a σ-algebra $\mathcal{E}$ containing $\mathcal{B}(E,\tau)$ and a unique measure μ on $(E,\mathcal{E})$ such that:

i) $\mu(K) < +\infty$ for any compact set K on E,

ii) For any $A \in \mathcal{E}$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

iii) For any $A \in \tau$ or $A \in \mathcal{E}$ with $\mu(A) < +\infty$,

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

iv) $(E,\mathcal{E},\mu)$ is complete

v) $\Lambda f = \int_E f d\mu$ for any $f \in C_c(E,\tau)$, where the integral on the right exists in $\mathbb{C}$.

Proof) The proof will consist of multiple steps. First, we construct the requisite measure μ on the discrete σ-algebra on E, after which we show that μ is countably additive on a sub σ-algebra of the discrete σ-algebra that contains the Borel sets. Finally, we show that μ satisfies the requirement $v)$.

The basic idea for the construction of μ proceeds as follows. For the integral of a function with respect to μ to be the value of the Λ given that function, the value of Λ given an indicator function must equal the measure of the associated set under the measure μ (this is in a heuristic sense; strictly speaking, Λ is not defined for indicator functions because they are not continuous). In order to make this hold, for any set $A \subset E$ we can define $\mu(A)$ as the supremum of Λf for any function $f \in C_c(E,\tau)$ that takes values between 0 and 1 and is 0 outside of A. Because Urysohn's lemma tells us it is possible to approximate indicator functions arbitrary well with continuous compactly supported functions on locally compact Hausdorff spaces, this way, $\mu(A)$ can equal the approximate value of Λ given the indicator function I_A.

Step 1: The Definition of μ

Defining μ for open sets

As per the idea introduced above, for any open set $A \subset E$, define

$$\mu(A) = \sup\{\Lambda f \mid f \prec A\}.$$

Note that any function $f \prec A$ is continuous and compactly supported, takes values in $[0,1]$, and whose support is a compact set contained in A. Note that $\mu(A)$ is bounded below by 0 because Λ is a positive linear functional and any $f \prec A$ is non-negative valued. Additionally, for any open sets $A_1, A_2 \in \tau$ such that $A_1 \subset A_2$, it follows that $f \prec A_2$ if $f \prec A_1$, and as such that

$$\mu(A_1) = \sup\{\Lambda f \mid f \prec A_1\} \leq \sup\{\Lambda f \mid f \prec A_2\} = \mu(A_2).$$

The fact that Λ is a linear transformation tells us that $\Lambda I_\emptyset = 0$, since $I_\emptyset$ is the unique additive identity on $C_c(E,\tau)$. Therefore,

$$\mu(\emptyset) = \sup\{\Lambda f \mid f \prec \emptyset\} = \Lambda I_\emptyset = 0,$$

since $f \prec \emptyset$ requires $f(x) = 0$ for any $x \in \emptyset^c = E$.

Defining μ for arbitrary subsets of E

Now define the function $\mu^o : 2^E \to [0,+\infty]$ as

$$\mu^o(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

for any $A \subset E$. Since $\mu(V)$ takes values in $[0,+\infty]$ for any open set V, it follows that $\mu^o(A)$ is bounded below by 0 as well. Therefore, μ^o does indeed map into $[0,+\infty]$.
In addition, for any $A_1 \subset A_2$, because $A_2 \subset V$ implies $A_1 \subset V$ as well,

$$\{\mu(V) \mid A_2 \subset V, V \in \tau\} \subset \{\mu(V) \mid A_1 \subset V, V \in \tau\}$$

and

$$\mu^o(A_1) = \inf\{\mu(V) \mid A_1 \subset V, V \in \tau\} \leq \inf\{\mu(V) \mid A_2 \subset V, V \in \tau\} = \mu^o(A_2).$$

This tells us that the function μ^o is montone.

The definition of μ^o shows us that the measure of an arbitrary subset A is determined as the approximation of the measure of open sets containing A, where the measure of some open set was defined above as the approximation of the value of Λ for the indicator of said open set. As such, $\mu^o(A)$ approximates the value of Λ for the indicator I_A.

Showing that the measure of an open set is well-defined

Before proceeding, note that the above definitions apparently provide two definitions of the measure of an open set $V \in \tau$, either as $\mu(V) = \sup\{\Lambda f \mid f \prec V\}$ or as $\mu^{\circ}(V) = \inf\{\mu(U) \mid V \subset U, U \in \tau\}$. However, because

$$\mu(V) \leq \mu(U)$$

for any $U \in \tau$ such that $V \subset U$, and V is itself an open set, we have the equality

$$\mu^{\circ}(V) = \inf\{\mu(U) \mid V \subset U, U \in \tau\} = \mu(V).$$

Therefore, the two definitions agree on the collection of all open sets, and we can denote $\mu^{\circ} = \mu$ on 2^E.

Step 2: μ as an Outer Measure

In this part, we will show that μ is an outer measure on E, that is, a function $\mu : 2^E \to [0,+\infty]$ such that $\mu(\emptyset) = 0$, $\mu(A) \leq \mu(B)$ if $A \subset B$, and which is countably subadditive. We already proved the first two properties, so it remains to see whether μ is countably subadditive.

Finite subadditivity of μ for open sets

To this end, we first prove that μ is finitely subadditive for open sets. Let $V_1, V_2 \in \tau$, and suppose $g \prec V_1 \cup V_2$. Since g is a continuous compactly supported function whose support K is contained in the open set $V_1 \cup V_2$, by the partition of unity theorem there exist functions h_1 and h_2 in $C_c(E,\tau)$ such that $h_1 \prec V_1$, $h_2 \prec V_2$ and $h_1(x) + h_2(x) = 1$ for any $x \in K$.
For $i = 1,2$, $h_i g$ is a continuous complex valued function because continuity is preserved across products. Additionally, $h_i(x)g(x) \in [0,1]$ because both h_i and g take values in $[0,1]$. From

$$\{h_i = 0\} \cup \{g = 0\} = \{h_i g = 0\},$$

we can see that

$$\overline{\{h_i g \neq 0\}} \subset \overline{\{h_i \neq 0\}} \cap K \subset V_1 \cap (V_1 \cup V_2) = V_1.$$

Because h_i is compactly supported, $\overline{\{h_i \neq 0\}} \cap K$ is compact, implying that $\overline{\{h_i g \neq 0\}}$ is also compact and contained in V_1. Therefore, $h_i g \prec V_i$.
The fact that $h_1 + h_2 = 1$ on K indicates that $h_1 g + h_2 g = g$ on K. If $x \notin K$, then $g(x) = 0$, meaning that $h_1(x)g(x) + h_2(x)g(x) = 0 = g(x)$. Thus, $h_1 g + h_2 g = g$ on E, and by the linearity of Λ and the definition of μ for open sets, we now have

$$\Lambda g = \Lambda(h_1 g) + \Lambda(h_2 g) \leq \mu(V_1) + \mu(V_2).$$

This holds for any $g \prec V_1 \cup V_2$, so we have

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2).$$

It follows by induction that

$$\mu\left(V_1 \cup \cdots \cup V_n\right) \leq \mu(V_1) + \cdots + \mu(V_n)$$

for any $n \in N_+$ and open sets $V_1, \cdots, V_n$.

Countable subadditivity of μ for arbitrary subsets of E

Now let $\{A_n\}_{n\in N_+}$ be an arbtirary collection of subsets of E, and denote $A = \bigcup_n A_n$. If $\mu(A_n) = +\infty$ for some $n \in N_+$, the inequality

$$\mu(A) \le \sum_{n=1}^{\infty} \mu(A_n)$$

holds trivially.
Suppose that $\mu(A_n) < +\infty$ for any $n \in N_+$. Then, for any $\varepsilon > 0$ and any $n \in N_+$, because

$$\mu(A_n) = \inf\{\mu(V) \mid A_n \subset V, V \in \tau\},$$

by the definition of the infimum there exists an open set V_n containing A_n such that

$$\mu(A_n) \le \mu(V_n) < \mu(A_n) + 2^{-n}\varepsilon.$$

Define $V = \bigcup_n V_n \in \tau$. For any $f \prec V$, because the support K of f is contained in V, $\{V_n\}_{n\in N_+}$ is an open cover of K and by compactness there exists an $m \in N_+$ such that

$$K \subset V_1 \cup \cdots \cup V_m.$$

f is thus a continuous compactly supported function taking values in $[0,1]$ whose support is contained in the open set $V_1 \cup \cdots \cup V_m$. By definition, $f \prec V_1 \cup \cdots \cup V_m$, and as such, by the definition of μ for open sets and the finite subadditivity result shown above,

$$\Lambda f \le \mu(V_1 \cup \cdots \cup V_m) \le \sum_{i=1}^{m} \mu(V_i) \le \sum_{n=1}^{\infty} \mu(V_n)$$

where the last inequality follows because μ is non-negative valued.
By how we chose $V_1, V_2, \cdots$,

$$\sum_{n=1}^{\infty} \mu(V_n) \le \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon \cdot \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon,$$

and we have

$$\mu(V) = \sup\{\Lambda f \mid f \prec V\} \le \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon.$$

Because $A \subset V$ (each $A_n \subset V_n$ by design), it follows that

$$\mu(A) \le \mu(V) \le \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon.$$

Finally, this holds for any $\varepsilon > 0$, which implies that

$$\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

We have therefore seen that μ is an outer measure on E.

Step 3: Value of μ for Compact Sets

We now study the value of $\mu(K)$ for any compact set K in E. Specifically, we will show that $\mu(K) < +\infty$ and

$$\mu(K) = \inf\{\Lambda f \mid K \prec f\}$$

for any compact K. Note that the set $\{\Lambda f \mid K \prec f\}$ is non-empty for any compact set K, since $K \subset E$, where E is open, and this ensures the existence of some continuous compactly supported function f such that $K \prec f \prec E$ by Urysohn's lemma. Any f such that $K \prec f$ is non-negative real valued, so by positivity, $\Lambda f \in [0, +\infty)$ as well. As such, the infimum on the right is well defined as a value in $[0, +\infty)$ by the least upper bound property of the real line.

Proving $\mu(K) \leq \inf\{\Lambda f \mid K \prec f\}$

First, choose any f such that $K \prec f$. Then, f is continous and compactly supported on E, takes values in $[0, 1]$, and equals 1 on K. f is non-negative real valued, so by the positivity of Λ, so is Λf.
For any $n \in N_+$, define

$$V_n = \{x \in E \mid f(x) > \frac{n-1}{n}\} = f^{-1}\left(\left(\frac{n-1}{n}, +\infty\right)\right).$$

Because f is a continuous real-valued function, $V_n \in \tau$, and we can see that $K \subset V_n$, since $f(x) = 1 > \frac{n-1}{n}$ for $x \in K$. For any $g \prec V_n$, because the support of g is contained in V_n, if $g(x) > 0$, then $x \in V_n$ and thus

$$f(x) > \frac{n-1}{n} \geq \frac{n-1}{n} g(x),$$

where the second equality follows from the fact that $g(x) \in [0, 1]$. If $x \notin V_n$, then $g(x) = 0$ and it is trivially true that $f(x) \geq \frac{n-1}{n} g(x)$. Thus, $\frac{n-1}{n} g \leq f$ on E, and as such, $f - \frac{n-1}{n} g \geq 0$ is a non-negative real valued function contained in $C_c(E, \tau)$ ($C_c(E, \tau)$ is a vector space containing f, g). It follows that

$$\begin{aligned}
\Lambda f &= \Lambda\left(f - \frac{n-1}{n} g\right) + \frac{n-1}{n} \cdot \Lambda g && \text{(Linearity of } \Lambda) \\
&\geq \frac{n-1}{n} \cdot \Lambda g, && \text{(Positivity of } \Lambda)
\end{aligned}$$

which implies that

$$\Lambda g \leq \frac{n}{n-1} \cdot \Lambda f.$$

This holds for any $g \prec V_n$, so

$$\mu(V_n) = \sup\{\Lambda g \mid g \prec V_n\} \leq \frac{n}{n-1} \cdot \Lambda f.$$

By the monotonicity of μ and the fact that $K \subset V_n$,

$$\mu(K) \leq \mu(V_n) \leq \frac{n}{n-1} \cdot \Lambda f;$$

this holds for any $n \in N_+$, so sending $n \to \infty$ yields

$$\mu(K) \leq \Lambda f.$$

Because $\Lambda f \in [0, +\infty)$, this also shows that $\mu(K) < +\infty$.
The above inequality holds for any f such that $K \prec f$, so

$$\mu(K) \leq \inf\{\Lambda f \mid K \prec f\}.$$

Proving $\mu(K) \geq \inf\{\Lambda f \mid K \prec f\}$

To see that the reverse inequality holds, choose any $\varepsilon > 0$; because

$$\mu(K) = \inf\{\mu(V) \mid K \subset V, V \in \tau\}$$

and $\mu(K) < +\infty$, by the definition of the infimum there exists a $V \in \tau$ such that $K \subset V$ and

$$\mu(K) \leq \mu(V) < \mu(K) + \varepsilon.$$

Then, by Urysohn's lemma, there exists a continuous compactly supported function g such that $K \prec g \prec V$, which implies by the definition of $\mu(V)$ that

$$\Lambda g \leq \mu(V) < \mu(K) + \varepsilon.$$

Therefore,

$$\inf\{\Lambda f \mid K \prec f\} \leq \Lambda g < \mu(K) + \varepsilon,$$

and because this holds for any $\varepsilon > 0$, the inequality

$$\inf\{\Lambda f \mid K \prec f\} \leq \mu(K)$$

holds. In conclusion, we have

$$\mu(K) = \inf\{\Lambda f \mid K \prec f\} \in [0, +\infty).$$

So far, we have studied the properties of the function $\mu : 2^E \to [0, +\infty]$ defined as

- For any open set V,

$$\mu(V) = \sup\{\Lambda g \mid f \prec V\}$$

- For any $A \subset E$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}.$$

We have subsequently found that:

- $\mu(\emptyset) = 0$
- $\mu(A) \leq \mu(B)$ if $A \subset B$
- For any collection $\{A_n\}_{n \in N_+} \in 2^E$,

$$\mu\left(\bigcup_n A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

so that μ is an outer measure on E, and that, for any compact set K,

$$\mu(K) = \inf\{\Lambda f \mid K \prec f\} \in [0, +\infty).$$

Step 4: Defining $\mathcal{E}_F$ and $\mathcal{E}$

Having defined μ as above and shown that it is an outer measure, any subset of E can be said to be outer measurable. Since we have allowd any subset of E to be outer measurable, one may naturally question whether arbitrary subsets of E are inner measurable, where inner measurability is defined in a similar manner but instead with supremums instead of infimums. We now set out to do just that, and it turns out that only some functions are inner measurable in our current construction.

The Collection $\mathcal{E}_F$

Define the subcollection $\mathcal{E}_F$ of 2^E as the sets $A \subset E$ such that $\mu(A) < +\infty$ and

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}.$$

In this sense, $\mathcal{E}_F$ is the collection of all "inner measurable" sets, in contrast to the outer measurability defined earlier. Note that $\mathcal{E}_F$ is nonempty, since $\emptyset \in \mathcal{E}_F$ ($\emptyset$ is the only compat subset of $\emptyset$, and $\mu(\emptyset) = 0$).
Furthermore, we can see that any subset $A \subset E$ with value 0 under μ is contained in $\mathcal{E}_F$. To see this, let K be a compact set such that $K \subset A$. Then, by the non-negativity and monotonicity of μ, $0 \leq \mu(K) \subset \mu(A) = 0$, which implies that $\mu(K) = 0$ and therefore that

$$\mu(A) = 0 = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}.$$

$\mathcal{E}_F$ also contains any compact K. This is because $\mu(K) \leq \mu(K')$ for any compact K' such that $K \subset K'$, K is compact itself, and $\mu(K) < +\infty$.

The measure of an open set $V \in \tau$ also satisfies

$$\mu(V) = \sup\{\mu(K) \mid K \subset V, K \text{ is compact}\}.$$

Let $V \in \tau$. Then, for any compact K such that $K \subset V$, by the monotonicity of μ we have $\mu(K) \leq \mu(V)$, so

$$\sup\{\mu(K) \mid K \subset V, K \text{ is compact}\} \leq \mu(V).$$

To show the reverse inequality, let $f \prec V$, and denote by K the compact support of f, which satisfies $K \subset V$. Then, for any open set U containing K, because f is a continuous compactly supported function taking values in $[0,1]$ and whose support is contained in U, $f \prec U$ and $\Lambda f \leq \mu(U)$ by the definition of μ for open sets. As such,

$$\Lambda f \leq \inf\{\mu(U) \mid K \subset U, U \in \tau\} = \mu(K)$$

by the definition of μ for arbitrary subsets of E. Therefore,

$$\Lambda f \leq \mu(K) \leq \sup\{\mu(K') \mid K' \subset V, K' \text{ is compact}\}.$$

The above inequality holds for any f such that $f \prec V$, so

$$\mu(V) = \sup\{\Lambda f \mid f \prec V\} \leq \sup\{\mu(K) \mid K \subset V, K \text{ is compact}\},$$

which shows us that

$$\mu(V) = \sup\{\mu(K) \mid K \subset V, K \text{ is compact}\}.$$

It follows now that any open $V \in \tau$ with finite value under μ is contained in $\mathcal{E}_F$.

The Collection $\mathcal{E}$

We now define the subcollection $\mathcal{E}$ of 2^E as follows:

$$\mathcal{E} = \{A \subset E \mid A \cap K \in \mathcal{E}_F \text{ for any compact } K\}.$$

It is immediately evident that $\emptyset, E \in \mathcal{E}$, since $\emptyset \cap K = \emptyset \in \mathcal{E}_F$ and $E \cap K = K \in \mathcal{E}_F$ for any compact K. Likewise, any $A \subset E$ such that $\mu(A) = 0$ is contained in $\mathcal{E}$, since, for any compact K, $A \cap K$ has value 0 under μ by monotonicity and is thus contained in $\mathcal{E}_F$.
Finally, any compact K is contained in $\mathcal{E}$ as well, since, for any compact K_1, $K \cap K_1$ is also compact with $\mu(K \cap K_1) \leq \mu(K) < +\infty$, which implies that $K \cap K_1 \in \mathcal{E}_F$.

Summarizing our findings, we have defined subcollections $\mathcal{E}_F$ and $\mathcal{E}$ of 2^E that contain the following sets:

- The empty set $\emptyset$
- Every compact set
- Any set with value 0 under μ.

Any open set with finite value under μ is also contained in $\mathcal{E}_F$, and $\mathcal{E}$ also contains the entire set E.

Step 5: The Countable Additivity of μ on $\mathcal{E}_F$

Finite Additivity for Compact Sets

We first show that, for any disjoint compact K_1 and K_2, $\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2)$. It follows from the countable subadditivity of μ that $\mu(K_1 + K_2) \leq \mu(K_1) + \mu(K_2)$. It remains to show that the reverse inequality holds true.

Denote $K = K_1 \cup K_2$, and choose any f such that $K \prec f$. Letting $V = K_2^c$, V is an open set because the compact set K_2 is closed, and becaues $K_1 \subset V$ from the fact that $K_1 \cap K_2 = \emptyset$, by Urysohn's lemma there exists a continuous compactly supported function g taking values in $[0,1]$ such that $K_1 \prec g \prec V$, so that $g(x) = 1$ for $x \in K_1$ and $g(x) = 0$ for any $x \notin V$, or equivalently $x \in K_2$.

Define $h_1 = fg$ and $h_2 = (1-g)f$. Both h_1 and h_2 take values in $[0,1]$, and their supports are contained in the support of f, which is compact; this implies that $h_1, h_2 \in C_c(E,\tau)$. Finally, if $x \in K_1$, then $f(x) = 1$ ($K_1 \subset K_1 \cup K_2 = K \prec f$) and $g(x) = 1$ ($K_1 \prec g$), so that $h_1(x) = 1$, and likewise, if $x \in K_2$, then $f(x) = 1$ ($K_2 \subset K_1 \cup K_2 = K \prec f$) and $g(x) = 0$ ($g \prec K_2^c$), so that $h_2(x) = 1$. As such, by definition $K_1 \prec h_1$ and $K_2 \prec h_2$, where $h_1 + h_2 = f$. By the linearity of Λ, we have

$$\Lambda h_1 + \Lambda h_2 = \Lambda f.$$

We derived a characterization for the value of μ for compact sets in step 3 that mirrored the definition of μ for open sets. According to this definition,

$$\mu(K_i) = \inf\{\Lambda k \mid K_i \prec k\} \leq \lambda h_i\}$$

for $i = 1,2$, so we have

$$\mu(K_1) + \mu(K_2) \leq \Lambda h_1 + \Lambda h_2 = \Lambda f.$$

This in turn holds for any f such that $K \prec f$, so it follows that

$$\mu(K_1) + \mu(K_2) \leq \mu(K).$$

We can now conclue that

$$\mu(K_1) + \mu(K_2) = \mu(K_1 \cup K_2)$$

for disjoint compact sets K_1, K_2. By induction, it follows that

$$\mu(K_1) + \cdots + \mu(K_n) = \mu(K_1 \cup \cdots \cup K_n)$$

for a finite collection $\{K_1, \cdots, K_n\}$ of disjoint compact sets.

Countable Additivity on $\mathcal{E}_F$

Now we move onto the main result. Let $\{A_n\}_{n\in N_+}$ be a sequence of disjoint sets in $\mathcal{E}_F$, and denote $A=\bigcup_n A_n$, which may or may not be in $\mathcal{E}_F$. Because μ is an outer measure, by countable subadditivity

$$\mu(A)\leq\sum_{n=1}^{\infty}\mu(A_n).$$

We need only show the reverse inequality holds using the properties that $A_1,A_2,\cdots$ are disjoint and in $\mathcal{E}_F$.

For any $\varepsilon>0$ and $n\in N_+$, because $\mu(A_n)<+\infty$ and

$$\mu(A_n)=\sup\{\mu(K)\mid K\subset A_n, K \text{ is compact}\}$$

by the fact that $A_n\in\mathcal{E}_F$, by the definition of the supremum we can see that there exists a compact set K_n such that $K_n\subset A_n$ and

$$\mu(A_n)-2^{-n}\varepsilon<\mu(K_n)\leq\mu(A_n).$$

Since $A_1,A_2,\cdots$ are disjoint and $K_n\subset A_n$ for each $n\in N_+$, it follows that the sequence $\{K_n\}_{n\in N_+}$ is also a disjoint colleciton of compact sets.

For any $n\in N_+$, by the result shown above,

$$\sum_{i=1}^{n}\mu(K_i)=\mu(K_1\cup\cdots\cup K_n).$$

Because $K_i\subset A_i\subset A$ for $1\leq i\leq n$, it follows that $K_1\cup\cdots\cup K_n\subset A$, and by the monotonicity of μ, we have

$$\sum_{i=1}^{n}\mu(K_i)=\mu(K_1\cup\cdots\cup K_n)\leq\mu(A).$$

This holds for any $n\in N_+$, so taking $n\to\infty$ on both sides,

$$\sum_{n=1}^{\infty}\mu(K_n)\leq\mu(A),$$

where the limit on the left hand side exists because $\{\mu(K_n)\}_{n\in N_+}$ is a sequence of non-negative reals.

Finally, from the way we chose each K_n, we can see that

$$\sum_{n=1}^{\infty}\mu(K_n)\geq\sum_{n=1}^{\infty}\mu(A_n)-\varepsilon\cdot\sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n=\sum_{n=1}^{\infty}\mu(A_n)-\varepsilon.$$

Therefore,

$$\sum_{n=1}^{\infty} \mu(A_n) - \varepsilon \leq \mu(A),$$

and because this holds for any $\varepsilon > 0$, we have

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu(A)$$

and by implication the countable additivity follows.

The Inclusion of $A = \bigcup_n A_n$ in $\mathcal{E}_F$

An important consequence of the above result is that $A \in \mathcal{E}_F$ if $\mu(A) < +\infty$.
In this case, for any $\varepsilon > 0$, since $\{\sum_{i=1}^{n} \mu(A_i)\}_{n \in N_+}$ is an increasing non-negative real valued sequence whose limit is $\mu(A)$, $\mu(A)$ is the supremum of that sequence and thus there exists an $m \in N_+$ such that

$$\mu(A) - \frac{\varepsilon}{2} < \sum_{i=1}^{m} \mu(A_i) \leq \mu(A).$$

As we noted above, for any $1 \leq i \leq m$ there exists a compact K_i such that

$$\mu(A_i) - 2^{-i}\frac{\varepsilon}{2} < \mu(K_i) \leq \mu(A_i),$$

so substituting this into the inequality above yields

$$\mu(A) - \frac{\varepsilon}{2} < \sum_{i=1}^{m} \mu(A_i) < \sum_{i=1}^{m} \mu(K_i) - \frac{\varepsilon}{2} \cdot \left(\sum_{i=1}^{m} 2^{-i} \right) \leq \sum_{i=1}^{m} \mu(K_i) - \frac{\varepsilon}{2}.$$

$K_1, \cdots, K_m$ are disjoint, so $\sum_{i=1}^{m} \mu(K_i) = \mu(K_1 \cup \cdots \cup K_m)$, and since $K_1 \cup \cdots \cup K_m$ is a compact set contained in A, we have shown that, for any $\varepsilon > 0$, there exists a compact $K \subset A$ such that

$$\mu(A) - \varepsilon < \mu(K).$$

Since $\mu(A)$ is obviously an upper bound of the set $\{\mu(K) \mid K \subset A, K \text{ is compact}\}$ by the monotonicity of μ, we have $\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$. This tells us that $A \in \mathcal{E}_F$ if $\mu(A) < +\infty$.

We have thus shown that:

- For any disjoint sequence $\{A_n\}_{n\in N_+}$ in $\mathcal{E}_F$,

$$\mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(A_n),$$

- For any disjoint sequence $\{A_n\}_{n\in N_+}$ in $\mathcal{E}_F$,

$$\bigcup_n A_n \in \mathcal{E}_F \text{ if } \mu\left(\bigcup_n A_n\right) < +\infty$$

Step 6: The Approximation Property of Sets in $\mathcal{E}_F$

This is a key property of the function μ, which will eventually allow us to approximate any measurable function with continuous functions.
Choose any $A \in \mathcal{E}_F$ and $\varepsilon > 0$. Then, because $\mu(A) < +\infty$ and A satisfies both

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\} \qquad \text{(by definition)}$$

and

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}, \qquad (\text{by } A \in \mathcal{E}_F)$$

there exist $V \in \tau$ and compact K such that $K \subset A \subset V$ and

$$\mu(A) - \frac{\varepsilon}{2} < \mu(K) \leq \mu(A) \leq \mu(V) < \mu(A) + \frac{\varepsilon}{2}.$$

Becasue V is an open set such that $\mu(V) < \mu(A) + \frac{\varepsilon}{2} < +\infty$, so $V \setminus K = V \cap K^c$ is an open set also with finite value under μ by monotonicity. K is in $\mathcal{E}_F$ by compactness and $V \setminus K$ and K are disjoint, so from the countable additivity result shown earlier, we now have

$$\mu(V) = \mu(V \setminus K) + \mu(K),$$

which implies

$$\mu(V \setminus K) = \mu(V) - \mu(K) < \left(\mu(A) + \frac{\varepsilon}{2}\right) - \left(\mu(A) - \frac{\varepsilon}{2}\right) = \varepsilon.$$

We have therefore seen that, for any $A \in \mathcal{E}_F$ and $\varepsilon > 0$, there exist an open set V and a compact set K such that $K \subset A \subset V$ and

$$\mu(V \setminus K) < \varepsilon.$$

This shows us that any set in $\mathcal{E}_F$ can be approximated arbitrarily closely above by an open set and below by a compact set.

Step 7: Closedness of $\mathcal{E}_F$ under Finite Set Operations

We will show in this step that, for any $A,B \in \mathcal{E}_F$, $A \cup B, A \cap B, A \setminus B$ are all contained in $\mathcal{E}_F$ by using the approximation proprety derived in step 6.

We first show that, for any $A_1, A_2 \in \mathcal{E}_F$, $A_1 \setminus A_2 \in \mathcal{E}_F$.
$A_1 \setminus A_2$ clearly has finite value under μ by monotonicity and the finiteness of A, and by the approximation property derived above, for any $\varepsilon > 0$ there exist compact K_i and open V_i such that $K_i \subset A_i \subset V_i$ and $\mu(V_i \setminus K_i) < \frac{\varepsilon}{2}$ for $i = 1,2$. The difference $A_1 \setminus A_2$ is then majorized as follows:

$$A_1 \setminus A_2 = A_1 \cap A_2^c \subset V_1 \cap K_2^c = V_1 \setminus K_2.$$

For any $x \in V_1 \setminus K_2$, $x \in V_1$ and $x \notin K_2$; we can now consider two distinct caes. If $x \in K_1$, then $x \in K_1$ and $x \notin K_2$, so that $x \in K_1 \setminus K_2$, and if $x \notin K_1$, then because $x \in V_1$, $x \in V_1 \setminus K_1$. This shows that x is either in $V_1 \setminus K_1$ or in $K_1 \setminus K_2$.
Likewise, for any $x \in K_1 \setminus K_2$, $x \in K_1$ and $x \notin K_2$. If $x \in V_2$, then because $x \notin K_2$, we have $x \in V_2 \setminus K_2$, while if $x \notin V_2$, then because $x \in K_1$, we have $x \in K_1 \setminus V_2$. It follows that

$$\begin{aligned} A_1 \setminus A_2 \subset V_1 \setminus K_2 &\subset (V_1 \setminus K_1) \cup (K_1 \setminus K_2) \\ &\subset (V_1 \setminus K_1) \cup (K_1 \setminus V_2) \cup (V_2 \setminus K_2). \end{aligned}$$

By the countably subadditivity of μ,

$$\mu(A_1 \setminus A_2) \leq \mu(V_1 \setminus K_1) + \mu(K_1 \setminus V_2) + \mu(V_2 \setminus K_2) < \mu(K_1 \setminus V_2) + \varepsilon.$$

Since $K_1 \setminus V_2 = K_1 \cap V_2^c$ is a closed subset of the compact set K_1, it is itself compact; since $K_1 \subset A_1$ and $V_2^c \subset A_2^c$, we can also see that $K_1 \setminus V_2 \subset A_1 \setminus A_2$. Therefore, $K_1 \setminus V_2$ is a compact subset of $A_1 \setminus A_2$, which tells us that

$$\mu(A_1 \setminus A_2) < \mu(K_1 \setminus V_2) + \varepsilon \leq \sup\{\mu(K) \mid K \subset A_1 \setminus A_2, K \text{ is compact}\} + \varepsilon.$$

$\sup\{\mu(K) \mid K \subset A_1 \setminus A_2, K \text{ is compact}\}$ is bounded above by $\mu(A_1 \setminus A_2) < +\infty$, and the above inequality holds for any $\varepsilon > 0$, so

$$\mu(A_1 \setminus A_2) = \sup\{\mu(K) \mid K \subset A_1 \setminus A_2, K \text{ is compact}\}.$$

This, together with the fact that $\mu(A_1 \setminus A_2) < +\infty$, tells us that $A_1 \setminus A_2 \in \mathcal{E}_F$ by definition.

Now let $A, B \in \mathcal{E}_F$. We have already seen that $A \setminus B \in \mathcal{E}_F$, and since

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

where $\mu(A \cup B) \leq \mu(A) + \mu(B) < +\infty$ and $B, A \setminus B \in \mathcal{E}_F$ are disjoint sets in $\mathcal{E}_F$, the final result

of step 5 shows us that $A \cup B \in \mathcal{E}_F$.
Finally, since

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

where $A, A \setminus B \in \mathcal{E}_F$, the above result tells us that $A \cap B \in \mathcal{E}_F$.

Step 8: $\mathcal{E}_F$ is the collection of all $A \in \mathcal{E}$ such that $\mu(A) < +\infty$

Let $A \in \mathcal{E}_F$. Then, for any compact K, because $K \in \mathcal{E}_F$ as well, $A \cap K \in \mathcal{E}_F$ because $\mathcal{E}_F$ is closed under finite intersections; this implies that $A \in \mathcal{E}$, and therefore that $\mathcal{E}_F \subset \mathcal{E}$.

Now suppose that $A \in \mathcal{E}$, and $\mu(A) < +\infty$. Then, because

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\},$$

there must exist a $V \in \tau$ such that $A \subset V$ and $\mu(V) < +\infty$, since otherwise, $\mu(A) = +\infty$. Since V is an open set with finite value under μ, it is contained in $\mathcal{E}_F$, and by the approximation property derived in step 6, for any $\varepsilon > 0$ there exist a compact set K and an open set V' such that $K \subset V \subset V'$ and $\mu(V' \setminus K) < \frac{\varepsilon}{2}$. By montonicity,

$$\mu(V \setminus K) \leq \mu(V' \setminus K) < \frac{\varepsilon}{2}.$$

Now, since $A \cap K \in \mathcal{E}_F$ by the compactness of K and the definition of A as a set in $\mathcal{E}$, the relationship

$$\mu(A \cap K) = \sup\{\mu(H) \mid H \subset A \cap K, H \text{ is compact}\}$$

holds by the definition of $\mathcal{E}_F$. Since $\mu(A \cap K) < +\infty$, there exists a compact set $H \subset A \cap K$ such that

$$\mu(A \cap K) - \frac{\varepsilon}{2} < \mu(H) \leq \mu(A \cap K),$$

which implies that

$$\begin{aligned}
\mu(A) =& \mu((A \cap K) \cup (A \setminus K)) \\
&\leq \mu((A \cap K) \cup (V \setminus K)) && \text{(Monotonicity; } A \subset V\text{)} \\
&= \mu(A \cap K) + \mu(V \setminus K) && \text{(Finite Additivity on } \mathcal{E}_F\text{; } A \cap, V \setminus K \text{ are disjoint sets in } \mathcal{E}_F\text{)} \\
&< \mu(H) + \varepsilon.
\end{aligned}$$

Because $H \subset A \cap K \subset A$, this shows us that

$$\mu(A) \leq \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\} + \varepsilon,$$

and because this holds for any $\varepsilon > 0$ and $\mu(A)$ is an upper bound of the set $\{\mu(K) \mid K \subset A, K \text{ is compact}\}$, we have $\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$.

This, alongside the assumption that $\mu(A) < +\infty$, shows us that

$$\mathcal{E}_F = \{A \in \mathcal{E} \mid \mu(A) < +\infty\}.$$

Step 9: $(E,\mathcal{E},\mu)$ is a Complete Measure Space and $\mathcal{B}(E,\tau) \subset \mathcal{E}$

$\mathcal{E}$ is a σ-algebra on E

We will first show that $\mathcal{E}$ is a σ-algebra on E.
Step 2 already shows that $\emptyset, E \in \mathcal{E}$.
Furthermore, for any $A \in \mathcal{E}$ and compact set K, since $A \cap K \in \mathcal{E}_F$ and

$$A^c \cap K = K \setminus A = K \setminus (A \cap K),$$

$A^c \cap K \in \mathcal{E}_F$ because $K \in \mathcal{E}_F$ by compactness. Therefore, $\mathcal{E}$ is closed under complements as well.

It remains to see that, for any countable collection $\{A_n\}_{n \in N_+}$ of sets in $\mathcal{E}$, their union $A = \bigcup_n A_n$ is in $\mathcal{E}$ as well. Fix a compact set K, and note that $A_n \cap K \in \mathcal{E}_F$ for any $n \in N_+$. From step 5, we saw that the countable union of a disjoint sequence of sets in $\mathcal{E}_F$ is also contained in $\mathcal{E}_F$ if its value under μ is finite. We will now exploit this finding.
Construct the sequence $\{B_n\}_{n \in N_+}$ of subsets of E as follows: $B_1 = A_1 \cap K$, and for any $n \geq 2$, define

$$B_n = (A_n \cap K) \setminus \left(\bigcup_{i=1}^{n-1} B_i \right).$$

By design, $\{B_n\}_{n \in N_+}$ is a disjoint sequence. $B_1 = A_1 \cap K \in \mathcal{E}_F$ by definition; now assume that $B_1, \cdots, B_n \in \mathcal{E}_F$ for some $n \geq 1$. then, because $\bigcup_{i=1}^n B_i \in \mathcal{E}_F$ ($\mathcal{E}_F$ is closed under finite unions) and $A_n \cap K \in \mathcal{E}_F$, $B_{n+1} \in \mathcal{E}_F$ because $\mathcal{E}_F$ is closed under differences. By induction, $\{B_n\}_{n \in N_+}$ is a disjoint sequence of sets in $\mathcal{E}_F$. Finally, the union of $\{B_1, \cdots, B_n\}$ is

$$\bigcup_{i=1}^{n} B_i = A_n \cap K$$

for any $n \geq 1$, so

$$\bigcup_n B_n = \bigcup_n (A_n \cap K) = A \cap K.$$

Since K is compact, it has finite value under μ, and by monotonicity, $\mu(A \cap K) \leq \mu(K) < +\infty$. Therefore, $A \cap K$ is the union of the disjoint sequence of sets $\{B_n\}_{n \in N_+}$ in $\mathcal{E}_F$ with $\mu(A \cap K) < +\infty$, which tells us that $A \cap K \in \mathcal{E}_F$ by step 5.
This holds for any compact K, so $A \in \mathcal{E}$ by definition, and we have shown that $\mathcal{E}$ is a σ-algebra on E.

Every Borel set is included in $\mathcal{E}$

For any open set V, the set $A = V^c$ is closed. For any compact K, $A \cap K$ is a closed subset of K and thus compact itself. Therefore, $A \cap K \in \mathcal{E}_F$, which tells us that $A \in \mathcal{E}$. This implies that

$V = A^c \in \mathcal{E}$ as well, since $\mathcal{E}$ is closed under complements. Therefore, $\tau \subset \mathcal{E}$, which tells us that $\mathcal{B}(E,\tau) \subset \mathcal{E}$ as well.

μ is a measure on $(E,\mathcal{E})$

We now show that μ is a measure on the measurable space $(E,\mathcal{E})$. Note initially that $\mu(\emptyset) = 0$; thus we need only prove the countable additivity of μ.

For any disjoint sequence of sets $\{A_n\}_{n\in N_+}$ in $\mathcal{E}$, denote $A = \bigcup_n A_n$. If $\mu(A_n) = +\infty$ for some $n \in N_+$, then by monotonicity, $\mu(A) = +\infty$ as well, meaning that the equality

$$\mu(A) = +\infty = \sum_{n=1}^{\infty} \mu(A_n)$$

holds trivially.

Now assume that $\mu(A_n) < +\infty$ for any $n \in N_+$. Then, because $\{A_n\}_{n\in N_+}$ is a disjoint sequence of sets in $\mathcal{E}_F$ with union A by step 8, by step 5 we have

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n).$$

$(E,\mathcal{E},\mu)$ is complete

It remains to be seen that $(E,\mathcal{E},\mu)$ is a complete measure space. To this end, let $A \in \mathcal{E}$ be a set with $\mu(A) = 0$, and let $N \subset A$. Then, by monotonicity $\mu(N) = 0$ as well, which tells us that $N \in \mathcal{E}$, as shown in step 4.

Step 10: $\Lambda f = \int_E f d\mu$ **for any** $f \in C_c(E,\tau)$

Any $f \in C_c(E,\tau)$ **is** μ**-integrable**

Let $f \in C_c(E,\tau)$ with compact support K. By continuity, f is a Borel measurable complex function. We will first show that f is μ-integrable.
Because $f(E)$ is a compact subset of the complex plane, by the Heine-Borel thoerem $f(E)$ is closed and bounded; thus, there exists an $M < +\infty$ such that $|f(x)| \leq M$ for any $x \in E$. In addition, since $f(x) = 0$ for any $x \notin K$, we can see that

$$\int_E |f| d\mu = \int_K |f| d\mu + \int_{K^c} |f| d\mu = \int_K |f| d\mu.$$

$|f| \leq M$ on K, so

$$\int_E |f| d\mu \leq M \cdot \mu(K) < +\infty,$$

where $\mu(K) < +\infty$ because K is compact. Therefore, f is a μ-integrable Borel measurable function, so that $f \in L^1(\mathcal{B}(E,\tau),\mu)$. Since $\mathcal{B}(E,\tau) \subset \mathcal{E}$, we can also say that $f \in L^1(\mathcal{E},\mu)$.

$\Lambda f = \int_E f d\mu$ **for real valued** $f \in C_c(E,\tau)$

Now we will show that $\Lambda f = \int_E f d\mu$.
First, assume that f is real-valued. Then, it suffices to show that $\Lambda f \leq \int_E f d\mu$, since the reverse inequality follows from the inequality

$$-\Lambda f = \Lambda(-f) \leq \int_E (-f) d\mu = -\int_E f d\mu,$$

which holds because $-f$ is a continuous compacty supported function on E if f is.

Since $f(E)$ is a compact subset of the real line, it is bounded, and as such there exist $a, b \in \mathbb{R}$ such that $f(E) \subset (a,b]$. For any $\varepsilon > 0$, choose $n \in N_+$ so that $\frac{b-a}{n} < \varepsilon$. Then, we define $\{y_0, \cdots, y_n\} \subset \mathbb{R}$ as

$$y_i = a + \frac{i}{n}(b-a)$$

for any $0 \leq i \leq n$. This shows us that

$$a = y_0 < \cdots < y_n = b$$

and that $y_i - y_{i-1} = \frac{b-a}{n} < \varepsilon$ for any $1 \leq i \leq n$.
Define

$$A_i = f^{-1}((y_{i-1}, y_i]) \cap K \text{ for any } 1 \leq i \leq n;$$

since $\{(y_0, y_1], \cdots, (y_{n-1}, y_n]\}$ is a partition of the range $(a, b]$ of f, and because they are all Borel sets in $\mathbb{R}$, and f is Borel measurable by continuity, each $f^{-1}((y_{i-1}, y_i]) \in \mathcal{B}(E, \tau)$. In addition,

$$\begin{aligned} A_1 \cup \cdots \cup A_n &= \left(\bigcup_{i=1}^{n} f^{-1}((y_{i-1}, y_i]) \right) \cap K \\ &= f^{-1}\left(\bigcup_{i=1}^{n} (y_{i-1}, y_i] \right) \cap K = f^{-1}((a, b]) \cap K = E \cap K = K, \end{aligned}$$

where $K \in \mathcal{B}(E, \tau)$ because K is compact, so $A_1, \cdots, A_n$ are Borel sets in E whose union equals the support K of f. Finally, because $\{(y_0, y_1], \cdots, (y_{n-1}, y_n]\}$ are disjoint, so are $A_1, \cdots, A_n$. Since $\mathcal{B}(E, \tau)$ is contained in $\mathcal{E}$, it follows that $A_1, \cdots, A_n$ are $\mathcal{E}$-measurable functions. Since $A_i \subset K$ for $1 \leq i \leq n$, by monotonicity $\mu(A_i) < +\infty$ for each i; this indicates, by step 8, that $A_1, \cdots, A_n$ are disjoint sets in $\mathcal{E}_F$.

By definition of the value of sets under μ, for $1 \leq i \leq n$ we have

$$\mu(A_i) = \inf\{\mu(V) \mid A_i \subset V, V \in \tau\}.$$

This means that there exists an open set $V' \in \tau$ such that $A \subset V'$ and

$$\mu(A_i) \leq \mu(V') < \mu(A_i) + \frac{\varepsilon}{n}.$$

Furthermore, because f is continuous and $(y_{i-1}, y_i + \varepsilon)$ is an open set in $\mathbb{R}$, the inverse image $V'' = f^{-1}((y_{i-1}, y_i + \varepsilon))$ is an open set in E that contains A_i, since

$$A_i = f^{-1}((y_{i-1}, y_i]) \subset f^{-1}((y_{i-1}, y_i + \varepsilon)) = V''.$$

Defining $V_i = V' \cap V'' \in \tau$, it follows that $A_i \subset V_i$ and

- $\mu(V_i) \leq \mu(V') < \mu(A_i) + \frac{\varepsilon}{n}$
- For any $x \in V_i$, since $x \in V''$ as well, $f(x) \in (y_{i-1}, y_i + \varepsilon)$, or equivalently, $y_{i-1} < f(x) < y_i + \varepsilon$. Furthermore, since $y_{i-1} < f(x) < y_i$ and $y_i - y_{i-1} < \varepsilon$, we have $0 < y_i - f(x) < y_i - y_{i-1} < \varepsilon$, which implies $f(x) > y_i - \varepsilon$.
  Therefore, if $x \in V_i$, then
  $$y_i - \varepsilon < f(x) < y_i + \varepsilon.$$

Because $A_i \subset V_i$ for each $1 \leq i \leq n$,

$$K = A_1 \cup \cdots \cup A_n \subset V_1 \cup \cdots \cup V_n.$$

Each V_i is open, while K is compact, so by the partition of unity theorem, there exist continuous compactly supported functions $h_1, \cdots, h_n$ such that $h_i \prec V_i$ for $1 \leq i \leq n$ and $K \prec \sum_{i=1}^{n} h_i$. Since

$\sum_{i=1}^n h_i = 1$ on K, while $f(x) = 0$ for any $x \in K^c$, we can see that $(\sum_{i=1}^n h_i) f = f$ on E, while $K \prec \sum_{i=1}^n h_i$ and the characterization of $\mu(K)$ derived in step 3 reveals that

$$\mu(K) \leq \Lambda\left(\sum_{i=1}^n h_i\right) = \sum_{i=1}^n \Lambda h_i.$$

For any $1 \leq i \leq n$, if $x \in V_i$, then $f(x) < y_i + \varepsilon$, so

$$h_i(x) f(x) < h_i(x)\left(y_i + \varepsilon\right),$$

while if $x \notin V_i$, then $h_i(x) = 0$ (since $h_i \prec V_i$), and

$$h_i(x) f(x) = 0 = h_i(x)\left(y_i + \varepsilon\right).$$

Therefore, $h_i f \leq (y_i + \varepsilon) h_i$ on E.
Finally, for any $1 \leq i \leq n$, because $h_i \prec V_i$, by the defition of the value of μ for open sets, we have

$$\Lambda h_i \leq \mu(V_i) < \mu(A_i) + \frac{\varepsilon}{n}.$$

Summarizing these findings, we can see that $h_1, \cdots, h_n$ are functions in $C_c(E, \tau)$ such that:

1) $(\sum_{i=1}^n h_i) f = f$ on E,

2) $\mu(K) \leq \sum_{i=1}^n \Lambda h_i$, and for any $1 \leq i \leq n$,

3) $h_i f \leq (y_i + \varepsilon) h_i$ on E and

4) $\Lambda h_i \leq \mu(A_i) + \frac{\varepsilon}{n}$.

Now we can see that

$$\begin{aligned}
\Lambda f = \Lambda\left(\sum_{i=1}^n h_i f\right) &= \sum_{i=1}^n \Lambda(h_i f) && \text{(fact 1) above)} \\
&\leq \sum_{i=1}^n (y_i + \varepsilon)\, \Lambda h_i && \text{(fact 2) above)} \\
&= \sum_{i=1}^n (|a| + y_i + \varepsilon)\, \Lambda h_i - |a| \cdot \sum_{i=1}^n \Lambda h_i \\
&\leq \sum_{i=1}^n (|a| + y_i + \varepsilon)\left(\mu(A_i) + \frac{\varepsilon}{n}\right) - |a| \cdot \mu(K) && \text{(facts 2) and 4) above)} \\
&= \sum_{i=1}^n (y_i - \varepsilon) \cdot \mu(A_i) + (|a| + 2\varepsilon) \cdot \left(\sum_{i=1}^n \mu(A_i)\right) \\
&\quad + \varepsilon \cdot (|a| + \varepsilon) + \frac{\varepsilon}{n}\left(\sum_{i=1}^n y_i\right) - |a| \cdot \mu(K).
\end{aligned}$$

Since $A_1, \cdots, A_n$ are disjoint sets in $\mathcal{E}$, by finite additivity

$$\mu(K) = \mu(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mu(A_i),$$

while $y_i \leq b$ for $1 \leq i \leq n$ implies that

$$\sum_{i=1}^{n} y_i \leq nb,$$

so we have

$$\begin{aligned} \Lambda f &\leq \sum_{i=1}^{n} (y_i - \varepsilon) \cdot \mu(A_i) + (|a| + 2\varepsilon) \cdot \mu(K) + \varepsilon \cdot (|a| + b + \varepsilon) - |a| \cdot \mu(K) \\ &= \sum_{i=1}^{n} (y_i - \varepsilon) \cdot \mu(A_i) + \varepsilon \cdot (2\mu(K) + |a| + b + \varepsilon). \end{aligned}$$

Define the function g as

$$g = \sum_{i=1}^{n} (y_i - \varepsilon) \cdot I_{A_i};$$

since $A_1, \cdots, A_n \in \mathcal{B}(E, \tau)$ and $y_1, \cdots, y_n \in \mathbb{R}$, g is the linear combination of $\mathcal{E}$-measurable real-valued functions and therefore $\mathcal{E}$-measurable itself.
Furthermore, for any $x \in E$, if $x \in K$, then $x \in A_i$ for a unique $1 \leq i \leq n$, so that

$$g(x) = y_i - \varepsilon < f(x).$$

On the other hand, if $x \notin K$, then $f(x) = g(x) = 0$, so $g \leq f$ on E. By the monotonicity of integration,

$$\sum_{i=1}^{n} (y_i - \varepsilon) \cdot \mu(A_i) = \int_E g d\mu \leq \int_E f d\mu,$$

and as such, we get the final inequality

$$\Lambda f \leq \int_E f d\mu + \varepsilon \cdot (2\mu(K) + |a| + b + \varepsilon).$$

This holds for any $\varepsilon > 0$, and the second term on the right hand side goes to 0 as ε does, so

$$\Lambda f \leq \int_E f d\mu.$$

It follows, as mentioned above, that

$$\Lambda f = \int_E f d\mu.$$

$\Lambda f = \int_E f d\mu$ for arbitrary $f \in C_c(E,\tau)$

Now suppose f is a general complex valued function in $C_c(E,\tau)$. Then, by the linearity of Λ and integration,

$$\begin{aligned}
\Lambda f &= \Lambda(Re(f)) + i \cdot \Lambda(Im(f)) \\
&= \int_E Re(f) d\mu + i \cdot \int_E Im(f) d\mu \\
&= \int_E f d\mu,
\end{aligned}$$

since $Re(f), Im(f) \in C_c(E,\tau)$ as well.

Step 11: The Uniqueness of μ

It remains to show that μ is the unique measure on $(E,\mathcal{E})$ satisfying claims $i)$ to $v)$ of the theorem. To this end, suppose that there exist two measures, μ_1 and μ_2, that satisfy $i)$ to $v)$. For any $A \in \mathcal{E}$, by $ii)$ we have

$$\mu_i(A) = \inf\{\mu_i(V) \mid A \subset V \mid V \in \tau\}$$

for $i = 1,2$, so it suffices to show that $\mu_1(V) = \mu_2(V)$ for any $V \in \tau$ to establish that $\mu_1 = \mu_2$ on $\mathcal{E}$. However, by $iii)$ it holds that

$$\mu_i(V) = \sup\{\mu_i(K) \mid K \subset V, K \text{ is compact}\}$$

for $i = 1,2$ and open $V \in \tau$. Therefore, if $\mu_1(K) = \mu_2(K) < +\infty$ for any compact set K, then $\mu_1(V) = \mu_2(V)$ for any open set V, which in turn implies that $\mu_1 = \mu_2$ on $\mathcal{E}$. We thus prove that $\mu_1(K) = \mu_2(K)$ for any compact set K.

Choose any compact set K. For any $i = 1,2$, because $\mu_i(K) < +\infty$ and

$$\mu_i(K) = \inf\{\mu_i(V) \mid K_i \subset V \mid V \in \tau\},$$

for any $\varepsilon > 0$ there exists an open set $V_i \in \tau$ such that $K_i \subset V_i$ and

$$\mu_i(K) \leq \mu_i(V_i) < \mu_i(K) + \varepsilon.$$

By Urysohn's lemma, therre exists a continuous compactly supported function f_i such that $K_i \prec f_i \prec V_i$.
If $x \in K$, then $f_i(x) = 1$, so that $I_K(x) = f_i(x)$, while if $x \notin K$, then $I_K(x) = 0 \leq f_i(x)$; it follows that $I_K \leq f_i$.
Similarly, if $x \notin V_i$, then $f_i(x) = 0 = I_{V_i}(x)$, while if $x \in V_i$, then $f_i(x) \leq 1 = I_{V_i}(x)$, which implies that $f_i \leq I_{V_i}$ on E.
Letting the subscript $-i$ denote 1 if $i = 2$ and 2 if $i = 1$, we have

$$\begin{aligned}
\mu_{-i}(K) = \int_E I_K d\mu_{-i} \leq \int_E f d\mu_{-i} & \qquad \text{(Monotonicity of Integration; } I_K \leq f \text{ on } E) \\
= \Lambda f = \int_E f d\mu_i & \qquad \text{(Claim } v); f \in C_c(E,\tau)) \\
\leq \int_E I_{V_i} d\mu_i & \qquad \text{(Monotonicity of Integration; } f \leq I_{V_i} \text{ on } E) \\
= \mu_i(V_i) < \mu_i(K) + \varepsilon. &
\end{aligned}$$

This holds for any $\varepsilon > 0$, so it follows that

$$\mu_{-i}(K) \leq \mu_i(K).$$

Finally, this holds for $i = 1$ or 2, so we have $\mu_1(K) = \mu_2(K)$ for any compat set K. As stated above, this implies that $\mu_1(A) = \mu_2(A)$ for any $A \in \mathcal{E}$, and as such the measure μ constructed across steps 1 to 10 is unique.
Q.E.D.

4.2 The Regularity of Measures

Let (E,τ) be a locally compact Hausdorff space. Any measure μ defined for sets in $\mathcal{B}(E,\tau)$ is called a Borel measure.
Now let $(E,\mathcal{E},\mu)$ be a measure space. μ is a Borel measure if and only if $\mathcal{B}(E,\tau) \subset \mathcal{E}$. If

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

for any set A in some collection $\mathbb{B}$, then we say that μ is outer regular for sets in $\mathbb{B}$.
Similarly, if μ satisfies

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

for any $A \in \mathbb{B}$, then μ is inner regular for sets in $\mathbb{B}$.
If μ is both outer and inner regular on $\mathcal{B}(E,\tau)$, then we say that μ is a regular measure.
If μ is outer regular for all Borel sets, inner regular for all open sets, and $\mu(K) < +\infty$ for any compact set K in E, then we say that μ is a Radon measure.

In light of the above definitions, we can see that, for any positive linear functional Λ defined on the set of all continuous compactly supported functions, the Riesz representation theorem proves the existence of a Radon measure μ that "represents" Λ in the sense that $\Lambda f = \int_E f d\mu$ for any continuous compactly supported function f. However, the theorem does not tell us that the measure μ constructed above is a regular measure. Indeed, it is possible to find a measure as constructed above that is not inner regular for every Borel set.

To ensure that μ is regular, we require additional restrictions on the topological properties of (E,τ). To this end, we define the following concepts.
Let (E,τ) be a topological space, and μ a measure on the measurable space $(E,\mathcal{E})$, where the σ-algebra $\mathcal{E}$ contains the Borel σ-algebra $\mathcal{B}(E,\tau)$.
For any $A \in \mathcal{E}$, we say that A is σ-compact if A is the countable union of compact sets.
Similarly, we say that $A \in \mathcal{E}$ has σ-finite measure under μ if there exists a countable measurable partition $\{A_n\}_{n \in N_+}$ of A such that $\mu(A_n) < +\infty$ for any $n \in N_+$. Note the similarities between this definition and that of σ-finite measures.

The following are simple properties of σ-compact sets and sets with σ-finite measure:

Lemma 4.2 Let (E,τ) be a locally compact Hausdorff space and $\mathcal{B}(E,\tau)$ the corresponding Borel σ-algebra. Let $(E,\mathcal{E},\mu)$ be a measure space with properties *i)* to *iv)* of the Riesz representation theorem: to state them for completeness, assume that

i) $\mu(K) < +\infty$ for any compact $K \subset E$

ii) Any $\mathcal{E}$-measurable set is outer regular; for any $A \in \mathcal{E}$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

iii) Any $\mathcal{E}$-measurable set with finite measure or any open set is inner regular; for any $A \in \mathcal{E}$ such that $\mu(A) < +\infty$ or $A \in \tau$,

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

iv) $(E,\mathcal{E},\mu)$ is complete, and $\mathcal{E}$ contains all Borel sets.

Then, the following hold true:

i) If $A \in \mathcal{E}$ is σ-compact, then it has σ-finite measure under μ.

ii) If $A \in \mathcal{E}$ has σ-finite measure under μ, then it is inner regular.

Proof) Suppose that $A \in \mathcal{E}$ is σ-compact. Then, there exists a sequence $\{K_n\}_{n\in N_+}$ of compact sets such that $A = \bigcup_n K_n$. Note that $\mu(K_n) < +\infty$ by property *i)* of the Riesz representation theorem.
Define the sequence $\{A_n\}_{n\in N_+}$ as $A_1 = K_1$ and

$$A_n = K_n \setminus \left(\bigcup_{i=1}^{n-1} K_i\right)$$

for $n \geq 2$. Then, each A_n is $\mathcal{E}$-measurable, and $\{A_n\}_{n\in N_+}$ is disjoint with union $\bigcup_n K_n = A$. Finally, by the monotonicity of measures,

$$\mu(A_n) \leq \mu(K_n) < +\infty$$

for any $n \in N_+$. This means that $\{A_n\}_{n\in N_+}$ is a measurable partition of A such that $\mu(A_n) < +\infty$ for each n, so by definition A has σ-finite measure under μ.

Now suppose that $A \in \mathcal{E}$ has σ-finite measure under μ. By definition, there exists a measurable partition $\{A_n\}_{n\in N_+}$ of A such that $\mu(A_n) < +\infty$ for any $n \in N_+$.
If $\mu(A) < +\infty$, then A is inner regular by property *iii)* of the Riesz representation theorem, so suppose that $\mu(A) = +\infty$. We need to show that

$$\sup\{\mu(K) \mid K \subset A, K \text{ is compact}\} = +\infty.$$

Choose any $M > 0$. From the countable additivity of measures, we have

$$+\infty = \mu(A) = \sum_{n=1}^{\infty} \mu(A_n).$$

Thus, there exists an $N \in N_+$ such that $\sum_{i=1}^{n} \mu(A_i) > M$ for any $n \geq N$. Defining

$$\varepsilon = \sum_{i=1}^{N} \mu(A_i) - M > 0,$$

because $\mu(A_i) < +\infty$ for any $1 \leq i \leq N$, $A_1, \cdots, A_N$ are inner regular sets, and as such, for any $1 \leq i \leq N$ there exists a compact set K_i such that $K_i \subset A_i$ and

$$\mu(A_i) - \frac{\varepsilon}{N} < \mu(K_i).$$

Because $A_1, \cdots, A_N$ are disjoint, so are $K_1, \cdots, K_N$, so that, by countable additivity,

$$\sum_{i=1}^{N} \mu(A_i) - \varepsilon < \sum_{i=1}^{N} \mu(K_i) = \mu(K)$$

for the compact set $K = K_1 \cup \cdots \cup K_N$. Since $K_i \subset A_i \subset A$ for each $1 \leq i \leq N$, $K \subset A$ and

$$M = \sum_{i=1}^{N} \mu(A_i) - \varepsilon < \mu(K) \leq \sup\{\mu(K') \mid K' \subset A, K' \text{ is compact}\}.$$

This holds for any $M > 0$, so that

$$\sup\{\mu(K) \mid K \subset A, K \text{ is compact}\} = +\infty = \mu(A).$$

Q.E.D.

Now we can show that, if the entire set E is σ-compact in addition to (E, τ) being a locally compact Hausdorff space, then the measure defined in the Riesz representation theorem is a regular measure in addition to being a Radon measure. Furthermore, σ-compactness allows us to extend the approximation property introduced in step 6 of the proof of the representation theorem to every measurable set, not just to those measurable sets with finite measure.

Theorem 4.3 Let (E,τ) be a locally compact Hausdorff space and $\mathcal{B}(E,\tau)$ the corresponding Borel σ-algebra. Let $(E,\mathcal{E},\mu)$ be a measure space with properties *i)* to *iv)* of the Riesz representation theorem:

i) $\mu(K) < +\infty$ for any compact $K \subset E$

ii) Any $\mathcal{E}$-measurable set is outer regular; for any $A \in \mathcal{E}$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

iii) Any $\mathcal{E}$-measurable set with finite measure or any open set is inner regular; for any $A \in \mathcal{E}$ such that $\mu(A) < +\infty$ or $A \in \tau$,

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

iv) $(E,\mathcal{E},\mu)$ is complete, and $\mathcal{E}$ contains all Borel sets.

If E is σ-compact, then the following hold true:

i) μ is a regular Borel measure.

ii) For any $A \in \mathcal{E}$ and $\varepsilon > 0$, there exist open and closed sets V and F such that $F \subset A \subset V$ and $\mu(V \setminus F) < \varepsilon$.

iii) For any $A \in \mathcal{E}$, there exist measurable sets F, V such that F is F_σ (the countable union of closed sets), V is G_δ (the countable intersection of open sets), $F \subset A \subset V$ and $\mu(V \setminus F) = 0$.

Proof) We will first show that μ is a regular measure. By σ-compactness, there exists a sequence $\{K_n\}_{n\in N_+}$ of compact sets such that $E = \bigcup_n K_n$. For any $A \in \mathcal{E}$,

$$A = A \cap E = \bigcup_n (A \cap K_n).$$

Defining $A_n = A \cap K_n$ for any $n \in N_+$, by monotonicity and the fact that $\mu(K_n) < +\infty$ by property *i)* of the representation theorem tells us that $\mu(A_n) < +\infty$.
Define the sequence $\{B_n\}_{n\in N_+}$ as $B_1 = A_1$ and

$$B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right)$$

for $n \geq 2$. Then, each B_n is $\mathcal{E}$-measurable, and $\{B_n\}_{n\in N_+}$ is disjoint with union $\bigcup_n A_n = A$. Finally, by the monotonicity of measures,

$$\mu(B_n) \leq \mu(A_n) < +\infty$$

for any $n \in N_+$. This means that $\{B_n\}_{n\in N_+}$ is a measurable partition of A such that $\mu(B_n) < +\infty$ for each n, so by definition A has σ-finite measure under μ.

By the preceding lemma, it now follows that A is inner regular. Since A is also outer regular by property $ii)$ of the representation theorem, and A was chosen arbitrarily from $\mathcal{E}$, we can see that μ is a regular Borel measure.

Now we will show that the approximation property holds.
Let $A \in \mathcal{E}$ and $\varepsilon > 0$. For any $n \in N_+$, note that $A \cap K_n$ is outer regular by property $ii)$ of the representation theorem and thus

$$\mu(A \cap K_n) = \inf\{\mu(V) \mid A \cap K_n \subset V, V \in \tau\}.$$

Because K_n and thus $A \cap K_n$ have finite measure under μ by property $i)$ of the representation theorem, by the definition of the infimum there exists an open set $V_n \in \tau$ containing $A \cap K_n$ such that

$$\mu(V_n) < \mu(A \cap K_n) + \frac{\varepsilon}{2^{n+2}}.$$

This implies that

$$\mu(V_n \setminus (A \cap K_n)) = \mu(V_n) - \mu(A \cap K_n) < \frac{\varepsilon}{2^{n+1}},$$

where the equality follows from finite additivity and the fact that $A \cap K_n \subset V_n$.
Defining $V = \bigcup_n V_n \in \tau$,

$$V \setminus A = \bigcup_n (V_n \setminus A) \subset \bigcup_n (V_n \setminus (A \cap K_n)),$$

so that, by monotonicity and countable subadditivity,

$$\mu(V \setminus A) \leq \sum_{n=1}^{\infty} \mu(V_n \setminus (A \cap K_n)) \leq \frac{\varepsilon}{4}.$$

The above result holds for A^c as well, so there exists an open set $U \in \tau$ such that $A^c \subset U$ and

$$\mu(U \setminus A^c) \leq \frac{\varepsilon}{4}.$$

Defining $F = U^c$, F is closed and $F \subset A$. Furthermore, since $U \setminus A^c = A \cap U = A \setminus F$, using the fact that

$$V \setminus F \subset (V \setminus A) \cup (A \setminus F),$$

we can see that

$$\mu(V \setminus F) \leq \mu(V \setminus A) + \mu(A \setminus F) \leq \frac{\varepsilon}{2} < \varepsilon.$$

To prove the final claim, choose any $A \in \mathcal{E}$. By the above result, for any $n \in N_+$ there exist open and closed sets V_n and F_n such that $F_n \subset A \subset V_n$ such that $\mu(V_n \setminus F_n) < \frac{1}{n}$. Defining $V = \bigcap_n V_n$ and $F = \bigcup_n F$, V is G_δ, F is F_σ, and

$$F = \bigcup_n F_n \subset A \subset \bigcap_n V_n = V.$$

For any $n \in N_+$,

$$V \setminus F \subset V_n \setminus F_n,$$

since $V \subset V_n$ and $F^c \subset F_n^c$. By monotonicity,

$$\mu(V \setminus F) \leq \mu(V_n \setminus F_n) < \frac{1}{n},$$

and because this holds for any $n \in N_+$, we have $\mu(V \setminus F) = 0$.
Q.E.D.

The following theorem furnishes sufficient conditions for an arbitrary Borel measure to be regular.

Theorem 4.4 Let (E,τ) be a locally compact Hausdorff space and $\mathcal{B}(E,\tau)$ the corresponding Borel σ-algebra. Let v be a measure on the measurable space $(E,\mathcal{B}(E,\tau))$, making v a Borel measure.
If every open set in E is σ-compact and every compact set in E has finite measure under v, then v is a regular Borel measure.

Proof) Let $f \in C_c(E,\tau)$ with compact support K. By continuity, f is a Borel measurable complex function. We will first show that f is v-integrable.
Because $f(E)$ is a compact subset of the complex plane, by the Heine-Borel thoerem $f(E)$ is closed and bounded; thus, there exists an $M < +\infty$ such that $|f(x)| \leq M$ for any $x \in E$. In addition, since $f(x) = 0$ for any $x \notin K$, we can see that

$$\int_E |f| dv = \int_K |f| dv + \int_{K^c} |f| dv = \int_K |f| dv.$$

$|f| \leq M$ on K, so

$$\int_E |f| dv \leq M \cdot v(K) < +\infty,$$

where $v(K) < +\infty$ because K is compact. Therefore, f is a v-integrable Borel measurable function, so that $f \in L^1(\mathcal{B}(E,\tau), v)$.

Now we can define the complex-valued function Λ on $C_c(E,\tau)$ as

$$\Lambda f = \int_E f dv$$

for any $f \in C_c(E,\tau)$. Λ is a linear functional because of the linearity of integration, and because the integral of a non-negative real-valued measurable function takes values in $[0,+\infty]$, Λ is positive as well.
By the Riesz representation theorem, because Λ is a linear functional on $C_c(E,\tau)$ and (E,τ) is locally compact and Hausdorff, there exists a σ-algebra $\mathcal{E}$ on E and a measure μ on $(E,\mathcal{E})$ satisfying the claims of the representation theorem.
Because E is itself an open set, it is σ-compact; by the previous theorem, μ is a regular measure. v will then be shown to be regular if we can just show that $v(A) = \mu(A)$ for any Borel set A. This is what we will show for the remainder of the proof.

We first show that $v(V) = \mu(V)$ for any open set $V \in \tau$. By the assumption of σ-compactness, there exists a sequence $\{K_n\}_{n \in N_+}$ of compact sets such that $V = \bigcup_n K_n$. For any $n \in N_+$, because $K_n \subset V$ and (E,τ) is a locally compact Hausdorff space, by

Urysohn's lemma there exists an $f_n \in C_c(E,\tau)$ such that $K_n \prec f_n \prec V$. Defining

$$g_n = \max(f_1, \cdots, f_n),$$

we can see that g_n, bing the maximum of continuous functions, is also continuous. Furthermore, because $\{f_1 = 0\} \cap \cdots \cap \{f_n = 0\} \subset \{g_n = 0\}$, we have

$$\overline{\{g_n \neq 0\}} \subset \overline{\{f_1 \neq 0\} \cup \cdots \cup \{f_n \neq 0\}} \subset \overline{\{f_1 \neq 0\}} \cup \cdots \cup \overline{\{f_n \neq 0\}},$$

where the set on the rightmost side is compact by the assumption that $f_1, \cdots, f_n$ are compactly supported. Thus, g_n is a continuous compactly supported function on E.
In addition, because $f_i(x) \in [0,1]$ for any $x \in E$, $g_n(x) \in [0,1]$ for any $x \in E$ as well, and the support of g_n is contained in V because $f_1, \cdots, f_n \prec V$. Finally, $f_i(x) = 1$ for any $x \in K_i$ and $1 \leq i \leq n$, so

$$g_n(x) = \max(f_1(x), \cdots, f_n(x)) = 1$$

for any $x \in K_1 \cup \cdots \cup K_n$.
We have so far constructed an increasing sequence $\{g_n\}_{n \in N_+}$ of continuous compactly supported functions taking values in $[0,1]$ such that, for any $n \in N_+$, $g_n(x) = 1$ for any $x \in K_1 \cup \cdots \cup K_n$ and $g_n(x) = 0$ for any $x \notin V$. Choose any $x \in E$; if $x \in V$, then because $V = \bigcup_n K_n$, $x \in K_N$ for some $N \in N_+$, which implies that

$$|I_V(x) - g_n(x)| = 0$$

for any $n \geq N$. On the other hand, if $x \notin V$, then $|I_V(x) - g_n(x)| = 0$ once again, so it follows by definition that $g_n \nearrow I_V$ pointwise.
$\{g_n\}_{n \in N_+} \subset C_c(E,\tau)$ is an increasing sequence of non-negative $\mathcal{E}$-measurable functions with pointwise limit I_V; it follows that

$$\begin{aligned}
v(V) = \int_E I_V dv &= \lim_{n\to\infty} \int_E g_n dv && \text{(The MCT)} \\
&= \lim_{n\to\infty} \Lambda g_n && \text{(Definition of } \Lambda\text{)} \\
&= \lim_{n\to\infty} \int_E g_n d\mu && \text{(The Riesz Representation Theorem)} \\
&= \int_E I_V d\mu = \mu(V). && \text{(The MCT)}
\end{aligned}$$

It remains to show that $\mu(A) = v(A)$ for any Borel set A.
First, choose any $A \in \mathcal{B}(E,\tau)$. Because E is σ-compact, by the approximation property shown above, for any $\varepsilon > 0$ there exist open and closed sets V and F such that $F \subset A \subset V$

such that $\mu(V \setminus F) < \varepsilon$. By finite additivity and the fact that $F \subset V$,

$$\mu(V) = \mu(F \cup (V \setminus F)) = \mu(F) + \mu(V \setminus F) \leq \mu(F) + \varepsilon,$$

where the inequality is not strict because $\mu(F)$ could be $+\infty$, and because V is open, $\mu(V) = v(V)$ and thus

$$v(V) \leq \mu(F) + \varepsilon.$$

Since $V \setminus F \in \tau$, we also have $v(V \setminus F) = \mu(V \setminus F) < \varepsilon$; it thus holds that

$$\mu(V) = v(V) = v(F) + v(V \setminus F) \leq v(F) + \varepsilon.$$

Therefore,

$$\begin{aligned} \mu(A) \leq \mu(V) \leq v(F) + \varepsilon \leq v(A) + \varepsilon \quad \text{and} \\ v(A) \leq v(V) \leq \mu(F) + \varepsilon \leq \mu(A) + \varepsilon \end{aligned}$$

by the monotonicity of measures.
If $\mu(A) = +\infty$ or $v(A) = +\infty$, then we immediately have $\mu(A) = v(A)$. If $\mu(A), v(A) < +\infty$, then the above implies that

$$|\mu(A) - v(A)| < \varepsilon,$$

and because this holds for any $\varepsilon > 0$, we have $\mu(A) = v(A)$.
Q.E.D.

4.3 The Lebesgue Measure

Using the results above, we can now construct the Lebesgue measure. The construction proceeds as follows: we first define the Riemann integral, and using it, a positive lienar functional on the euclidean k-space $\mathbb{R}^k$. Then, using the Riesz representation theorem, we claim that there exists a complete measure space such that the integral of any compactly supported continuous function on $\mathbb{R}^k$ with respect to the constructed measure equals its Riemann integral. This constructed measure will then be referred to as the Lebesgue measure on $\mathbb{R}^k$ and the integral with respect to this measure the Lebesgue integral.

We now formalize the above intuition, by first introducing some important concepts.
A k-cell on $\mathbb{R}^k$ is defined as any set of the form

$$W = (a_1, b_1) \times \cdots \times (a_k, b_k)$$

where each $a_i, b_i \in \mathbb{R}$, and each open interval can be taken to be a closed or half-open interval. The volume of a k-cell W is defined as

$$\text{vol}(W) = \prod_{i=1}^{k} (b_i - a_i).$$

Given a $k \times k$ matrix T and any point $x \in \mathbb{R}^k$, we define the translation and rotation of a set $A \subset \mathbb{R}^k$ by x and T as

$$TA + x = \{Ty + x \mid y \in \mathbb{R}^k\}.$$

Finally, given a $\delta > 0$ and $x \in \mathbb{R}^k$, the δ-box with corner at x is defined as

$$Q(x;\delta) = [x_1, x_1 + \delta) \times \cdots [x_k, x_k + \delta).$$

For any $n \in N_+$, let P_n be the set of all dyadic rationals with denominator 2^n and Ω_n the set of all 2^{-n}-boxes with corner at a point in P_n, that is,

$$\Omega_n = \{Q(x; 2^{-n}) \mid x \in P_n\}.$$

We can then show that the boxes in Ω_n have the following properties:

Lemma 4.5 The following hold true:

i) For any $n \in N_+$, any $x \in \mathbb{R}^k$ lies in one and only one box in Ω_n.

ii) If $r, n \in N_+$, $n < r$ and $Q \in \Omega_r$, $Q' \in \Omega_n$, then $Q \subset Q'$ or $Q \cap Q' = \emptyset$.

iii) For any $r \in N_+$, if $Q \in \Omega_r$, then $\mathrm{vol}(Q) = 2^{-rk}$. Furthermore, if $n \in N_+$ and $n > r$, then P_n has exactly $2^{(n-r)k}$ points in any $Q \in \Omega_r$.

iv) Any nonempty open set in $\mathbb{R}^k$ can be expressed as the countable union of disjoint boxes in $\bigcup_n \Omega_n$.

Proof) i) For any $x \in \mathbb{R}^k$ and $1 \le i \le k$, letting $m_i = \lfloor x_i \cdot 2^n \rfloor \in \mathbb{Z}$, we can see that

$$\frac{m_i}{2^n} \le x_i < \frac{m_i + 1}{2^n}.$$

Letting

$$y = \left[\frac{m_1}{2^n}, \frac{m_1+1}{2^n}\right) \times \cdots \times \left[\frac{m_k}{2^n}, \frac{m_k+1}{2^n}\right) \in P_n,$$

this implies that $x \in Q(y; 2^{-n}) \in \Omega_n$.
Now suppose that $x \in Q(y'; 2^{-n})$ for some $y' \in P_n$ such that $y \neq y'$. Then, there exists a $1 \le i \le k$ such that $y_i \neq y'_i$, or equivalently,

$$y'_i = \frac{z_i}{2^n}$$

for some $z_i \in \mathbb{Z}$ such that $z_i \neq m_i$. Assuming that $m_i < z_i$ without loss of generality, it follows that

$$x_i < \frac{m_i + 1}{2^n} \le \frac{z_i}{2^n},$$

so that $x_i \neq [y'_i, y'_i + 2^{-n})$ and thus $x \neq Q(y'; 2^{-n})$. Therefore, $Q(y; 2^{-n})$ is the only box in Ω_n in which x is contained.

ii) Choose any $r, n \in N_+$ such that $n < r$, and let $Q = Q(x; 2^{-r}) \in \Omega_r$, $Q' = Q(x'; 2^{-n}) \in \Omega_n$ for some $x \in P_r$ and $x' \in P_n$. Suppose that $Q \cap Q' \neq \emptyset$. Letting $y \in Q \cap Q'$, by definition

$$x_i = \frac{m_i}{2^r} \le y_i < x_i + 2^{-r} = \frac{m_i + 1}{2^r} \quad \text{and}$$
$$x'_i = \frac{z_i}{2^n} \le y_i < x'_i + 2^{-n} = \frac{z_i + 1}{2^n}$$

for any $1 \le i \le k$. Because $2^n < 2^r$, 2^{r-n} is an integer and, for any $1 \le i \le k$,

$$\frac{z_i \cdot 2^{r-n}}{2^r} \le y_i < \frac{(z_i + 1) \cdot 2^{r-n}}{2^r},$$

which implies that

$$z_i \cdot 2^{r-n} \leq m_i \quad \text{and} \quad m_i + 1 \leq (z_i + 1) \cdot 2^{r-n}.$$

Therefore, for $1 \leq i \leq k$,

$$x_i' = \frac{z_i}{2^n} \leq \frac{m_i}{2^r} = x_i < x_i + 2^{-r} \leq \frac{z_i + 1}{2^n} = x_i' + 2^{-n}$$

so that

$$[x_i, x_i + 2^{-r}) \subset [x_i', x_i' + 2^{-n}).$$

By implication, $Q \subset Q'$.

iii) For any $r \in N_+$ and $Q = Q(x; 2^{-r}) \in \Omega_r$, because

$$Q = \left[x_1, x_1 + 2^{-r}\right) \times \cdots \times \left[x_k, x_k + 2^{-r}\right)$$

$$\mathrm{vol}(Q) = \prod_{i=1}^{k} \left((x_i + 2^{-r}) - x_i\right) = \prod_{i=1}^{k} 2^{-r} = 2^{-rk}.$$

Now let $n \in N_+$ and suppose that $n > r$. Let $y \in P_n$ be contained in $Q = Q(x; 2^{-r})$. Then, for any $1 \leq i \leq k$,

$$x_i = \frac{m_i}{2^r} \leq y_i = \frac{z_i}{2^n} < \frac{m_i + 1}{2^r} = x_i + 2^{-r}.$$

It follows that

$$m_i \cdot 2^{n-r} \leq z_i < m_i \cdot 2^{n-r} + 2^{n-r},$$

so z_i must be equal to one of the 2^{n-r} integers

$$m_i \cdot 2^{n-r}, m_i \cdot 2^{n-r} + 1, \cdots, m_i \cdot 2^{n-r} + 2^{n-r} - 1.$$

This holds for any $1 \leq i \leq k$, so y must be one of $2^{(n-r)k}$ points in P_n, which proves the claim.

iv) Let $V \in \mathbb{R}^k$ be an open set, and choose any $x \in V$. Then, there exists a $\varepsilon > 0$ such that the ε-ball $B(x, \varepsilon)$ is contained in V. Letting $n \in N_+$ be chosen so that $2^{-n} < \frac{\varepsilon}{\sqrt{k}}$, it follows that

$$(x_1 - 2^{-n}, x_1 + 2^{-n}) \times \cdots \times (x_k - 2^{-n}, x_k + 2^{-n}) \subset V.$$

Letting $Q(y;2^{-n})$ be the box in Ω_n in which x is contained,

$$y_i \leq x_i < y_i + 2^{-n}$$

for any $1 \leq i \leq k$, which implies that

$$x_i - 2^{-n} < y_i < y_i + 2^{-n} \leq x_i + 2^{-n}$$

and therefore that

$$x \in Q(y;2^{-n}) = [y_1, y_1 + 2^{-n}) \times \cdots [y_k, y_k + 2^{-n}) \subset V.$$

It follows that

$$x \in \bigcup_n \bigcup_{Q \in \Omega_n, Q \subset V} Q(y;2^{-n}),$$

and as such that

$$V \subset \bigcup_n \bigcup_{Q \in \Omega_n, Q \subset V} Q(y;2^{-n}).$$

The converse is clearly true, so we have

$$V = \bigcup_n \bigcup_{Q \in \Omega_n, Q \subset V} Q(y;2^{-n}).$$

Note that, because each Ω_n is a countable collection of boxes, the above union is a countable union of boxes in $\Omega_1, \Omega_2, \cdots$ as well.

Let $\{W_n\}_{n \in N_+}$ be the countable collection of sets in $\Omega_1 \cup \Omega_2 \cup \cdots$ such that $V = \bigcup_n W_n$, ordered so that any set contained in Ω_n precedes those contained in Ω_{n+1}. Then, define $B_1 = W_1$ and

$$B_n = W_n \setminus \bigcup_{i=1}^{n-1} W_i.$$

Now $\{B_n\}_{n \in N_+}$ is a collection of disjoint measurable sets whose union is V. Additionally, because Ω_{n+1} is a collection of disjoint boxes that are either contained in some box in Ω_n or disjoint from it, $\{B_n\}_{n \in N_+}$ is a collection of disjoint boxes in $\Omega_1 \cup \Omega_2 \cup \cdots$ whose union is V.

Q.E.D.

The following is another useful lemma that concerns the continuity of continuous compactly supported functions.

Lemma 4.6 Let $f:\mathbb{R}^k \to \mathbb{C}$ be a continuous compactly supported function on the euclidean k-space $\mathbb{R}^k$. Then, f is uniformly continuous on $\mathbb{R}^k$.

Proof) Denote the euclidean metric on $\mathbb{R}^k$ by d_k. Let $K = \overline{\{f=0\}}$ be the compact support of f. By the Heine-Borel theorem, K is closed and bounded, so there exists an $M>0$ such that

$$K \subset [-M,M]^k.$$

Choose any $\epsilon > 0$. For any $x \in \mathbb{R}^k$, since f is continuous at x, there exists a $\delta_x \in (0,1)$ such that

$$|f(x)-f(y)| < \frac{\epsilon}{2}$$

for any $y \in \mathbb{R}^k$ such that $|x-y| < \delta_x$. The collection $\{B_{d_k}(x,\delta_x/2)\}_{x\in[-M-1,M+1]^k}$ is an open cover of the compact set $[-M-1,M+1]^k$, which shows us that there exists a finite set of points $\{x_1,\cdots,x_n\} \subset [-M-1,M+1]^k$ such that

$$[-M-1,M+1]^k \subset \bigcup_{i=1}^{n} B_{d_k}(x_i,\delta_{x_i}/2).$$

Define $\delta = \frac{1}{2}\min(\delta_{x_1},\cdots,\delta_{x_n}) \in (0,1)$. Choose any $x,y \in \mathbb{R}^k$ such that $|x-y| < \delta$, and consider three cases:

- $x,y \in K$
  In this case, both x and y are contained in $[-M-1,M+1]^k$. Letting $x \in B_{d_k}(x_i,\delta_{x_i}/2)$ for some $1 \le i \le n$, note that

  $$|x_i - y| \le |x_i - x| + |x-y| < \delta + \frac{\delta_{x_i}}{2} \le \delta_{x_i},$$

  so that $y \in B_{d_k}(x_i,\delta_{x_i}/2)$ as well. Therefore,

  $$|f(x)-f(y)| \le |f(x)-f(x_i)| + |f(x_i)-f(y)| < \epsilon.$$

- $x \in K, y \notin K$
  In this case, because

  $$|x-y| < \delta < 1$$

  and $x \in K \subset [-M,M]^k$, it follows that $x,y \in [-M-1,M+1]^k$. Therefore, the same line of reasoning as above implies that

$$|f(x) - f(y)| < \epsilon.$$

A symmetric argument shows that this inequality holds when $x \notin K$ and $y \in K$.

- $x, y \notin K$
  In this case, since $f(x) = f(y) = 0$,

  $$|f(x) - f(y)| = 0 < \epsilon$$

  trivially.

In any case, $|x - y| < \delta$ implies that $|f(x) - f(y)| < \epsilon$. Since such a $\delta > 0$ exists for any $\epsilon > 0$, by definition f is uniformly continuous on $\mathbb{R}^k$.

Q.E.D.

We are now ready to state the main result of this section.

Theorem 4.7 (Construction of the Lebesgue Measure)
There exists a σ-algebra $\mathcal{L}$ on $\mathbb{R}^k$ and a unique measure λ on $(\mathbb{R}^k, \mathcal{L})$ that satisfies the following properties:

i) $\mathcal{L}$ contains every Borel set.

ii) $(\mathbb{R}^k, \mathcal{L}, \lambda)$ is a complete measure space.

iii) $\lambda(K) < +\infty$ for any compact set $K \subset \mathbb{R}^k$.

iv) λ is a regular Borel measure.

v) For any $A \in \mathcal{L}$ and $\varepsilon > 0$, there exists a closed set F and an open set V such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) < \varepsilon.$$

vi) For any $A \subset \mathbb{R}^k$, $A \in \mathcal{L}$ if and only if there exist an F that is F_σ and a V that is G_δ such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) = 0.$$

vii) $\lambda(W) = \mathrm{vol}(W)$ for any k-cell W.

viii) λ is translation-invariant, that is, for any $A \in \mathcal{L}$ and $x \in \mathbb{R}^k$,

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

ix) If μ is a measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ that is translation-invariant and assigns finite measure to every compact set in $\mathbb{R}^k$, then there exists a $c > 0$ such that

$$\mu(A) = c \cdot \lambda(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$.

x) For any $k \times k$ matrix T, there exists a $\Delta(T) \in \mathbb{R}_+$ such that

$$\lambda(TA) = \Delta(T) \cdot \lambda(A)$$

for any $A \in \mathcal{L}$. In particular, if T is noninvertible, then $\Delta(T) = 0$.

Proof) We sequentially prove each claim in the theorem. Let τ_k be the standard euclidean topology on $\mathbb{R}^k$; note that $(\mathbb{R}^k, \tau_k)$ is a locally compact Hausdorff space in which every open set is σ-compact.

Step 1: Constructing the Riemann Integral

For any $n \in N_+$, define Λ_n as

$$\Lambda_n f = 2^{-kn} \cdot \sum_{x \in P_n} f(x)$$

for any complex valued function f on $\mathbb{R}^k$ with compact support (f is not necessarily continuous). The fact that f has compact support ensures that $\Lambda_n f \in \mathbb{C}$ and does not diverge, since only a finite number of $f(x)$ are non-zero.
Note that Λ_n is a positive linear functional; for any $a \in \mathbb{C}$ and complex functions f, g on $\mathbb{R}^k$ with compact support,

$$\begin{aligned}\Lambda_n(af + g) &= 2^{-kn} \cdot \sum_{x \in P_n} (a \cdot f(x) + g(x)) \\ &= a \cdot 2^{-kn} \sum_{x \in P_n} f(x) + 2^{-kn} \sum_{x \in P_n} g(x) = a \cdot \Lambda_n f + \Lambda_n g,\end{aligned}$$

while if $f \geq 0$, then

$$\Lambda_n f = 2^{-kn} \cdot \sum_{x \in P_n} f(x) \geq 0$$

as well. By implication, Λ_n is monotonic; that is, for any real valued f, g such that $f \leq g$,

$$\Lambda_n g = \Lambda_n f + \Lambda_n(g - f) \geq \Lambda_n f,$$

since $g - f \geq 0$ and thus $\Lambda_n(g - f) \geq 0$.

Denote by $C_c(\mathbb{R}^k)$ the collection of all continuous compactly supported functions on $\mathbb{R}^k$. Let $f \in C_c(\mathbb{R}^k)$ be a real valued function with support $K = \overline{\{f \neq 0\}}$, which is compact. By the preceding result, f is uniformly continuous on $\mathbb{R}^k$, so that, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$ for any $x, y \in \mathbb{R}^k$. Letting $N \in N_+$ be chosen so that $2^{-kN} < \delta$, define $g, h : \mathbb{R}^k \to \mathbb{C}$ as

$$g(x) = \min_{y \in Q} f(x) \quad \text{and} \quad h(x) = \max_{y \in Q} f(x)$$

for any $x \in Q$ for some $Q \in \Omega_N$. Then, g, h are constant on each box in Ω_N, $g \leq f \leq h$ on $\mathbb{R}^k$, and because the distance between any two points in some $Q \in \Omega_N$ is never greater than $2^{-kN} < \delta$, by the uniform continuity result above we have

$$h(x) - g(x) = \max_{y \in Q} f(x) - \min_{y \in Q} f(x) < \varepsilon$$

for any $x \in Q$, and by extension $h - g < \varepsilon$ on $\mathbb{R}^k$. Finally, because $f(x) = 0$ for any

$x \notin K$, we can see that $g(x) = h(x) = 0$ for any $x \notin K$ as well, so that K contains the support of g and h. In other words, g and h are compactly supported functions.

For any $n \geq N$, we can now see that

$$\begin{aligned}\Lambda_N g = 2^{-kN} \cdot \sum_{x \in P_N} g(x) &= \sum_{x \in P_N} g(x) \cdot \operatorname{vol}\left(Q(x; 2^{-N})\right) \\ &= \sum_{x \in P_n} 2^{-(n-N)k} g(x) \cdot \operatorname{vol}\left(Q(x; 2^{-n})\right) 2^{(n-N)k} \\ &= \sum_{x \in P_n} g(x) \cdot \operatorname{vol}\left(Q(x; 2^{-n})\right) = \Lambda_n g,\end{aligned}$$

where the third equality follows because there are exactly $2^{(n-N)k}$ points in P_n contained in each box in Ω_N, on which g is constant, so that $\Lambda_n g$ duplicates each term in $\Lambda_N g$ $2^{(n-N)k}$ times. Λ_n is monotonic, so

$$\Lambda_N g = \Lambda_n g \leq \Lambda_n f \leq \Lambda_n h = \Lambda_N h.$$

It now follows that

$$\begin{aligned}\limsup_{n \to \infty} \Lambda_n f - \liminf_{n \to \infty} \Lambda_n f \leq \Lambda_N (h - g) &= 2^{-kN} \cdot \sum_{x \in P_N} (h(x) - g(x)) \\ &\leq \varepsilon \cdot \sum_{Q \in \Omega_N, Q \subset W} \operatorname{vol}(Q) \leq \varepsilon \cdot \operatorname{vol}(W).\end{aligned}$$

This holds for any $\varepsilon > 0$, so we have

$$\limsup_{n \to \infty} \Lambda_n f = \liminf_{n \to \infty} \Lambda_n f = \lim_{n \to \infty} \Lambda_n f.$$

Denoting the above limit by Λf, we have shown that

$$\Lambda f = \lim_{n \to \infty} \Lambda_n f$$

exists for any real valued $f \in C_c(\mathbb{R}^k)$. This result can be easily extended to complex-valued $f \in C_c(\mathbb{R}^k)$ by the linearity of Λ_n for any $n \in N_+$. Λf can be interpreted as the usual Riemann integral for any $f \in C_c(\mathbb{R}^k)$.

Step 2: The Lebesgue Measure and Lebesgue Measurable Sets

Above, we have constructed a function $\Lambda : C_c(\mathbb{R}^k) \to \mathbb{C}$ such that

$$\Lambda(af+g) = \lim_{n\to\infty} \Lambda_n(af+g) = a \cdot \lim_{n\to\infty} \Lambda_n f + \lim_{n\to\infty} \Lambda_n g = a \cdot \Lambda f + \Lambda g$$

for any $a \in \mathbb{C}$ and $f, g \in C_c(\mathbb{R}^k)$ by the linearity of each Λ_n, and

$$\Lambda f = \lim_{n\to\infty} \Lambda_n f \geq 0$$

for any $f \in C_c(\mathbb{R}^k)$ such that $f \geq 0$ due to the positivity of each Λ_n. In other words, Λ is a positive linear functional on $C_c(\mathbb{R}^k)$, where $(\mathbb{R}^k, \tau_k)$ is a locally compact Hausdorff space where $\mathbb{R}^k$ is σ-compact.

Therefore, by the Riesz Representation Theorem and theorem 4.3, there exists a σ-algebra $\mathcal{L}$ on $\mathbb{R}^k$ and a unique measure λ on $(\mathbb{R}^k, \mathcal{L})$ such that:

- $\mathcal{B}(\mathbb{R}^k) \subset \mathcal{L}$, that is, every Borel set is contained in $\mathcal{L}$

- $(\mathbb{R}^k, \mathcal{L}, \lambda)$ is a complete measure space

- $\lambda(K) < +\infty$ for any compact set $K \subset \mathbb{R}^k$

- λ is a regular Borel measure

- For any $A \in \mathcal{L}$ and $\varepsilon > 0$, there exists a closed set F and an open set V such that $F \subset A \subset V$ and

  $$\lambda(V \setminus F) < \varepsilon$$

- For any $A \in \mathcal{L}$, there exists an F that is F_σ (the countable union of closed sets) and a V that is G_δ (the countable intersection of open sets) such that $F \subset A \subset V$ and

  $$\lambda(V \setminus F) = 0$$

– For any $f \in C_c(\mathbb{R}^k)$,

$$\Lambda f = \int_{\mathbb{R}^k} f d\lambda.$$

Conversely, if, for some $A \subset \mathbb{R}^k$, if there exists an F that is F_σ and a V that is G_δ such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) = 0,$$

then $A \setminus F \subset V \setminus F$ is a negligible set; becuase $(\mathbb{R}^k, \mathcal{L}, \lambda)$ is complete and $V \setminus F \in \mathcal{B}(\mathbb{R}^k) \subset \mathcal{L}$, $A \setminus F$ is also $\mathcal{L}$-measurable, so that

$$A = F \cup (A \setminus F) \in \mathcal{L}.$$

It follows that $A \in \mathcal{L}$ if and only if there exist an F that is F_σ and a V that is V_δ such that $F \subset A \subset V$ and $\lambda(V \setminus F) = 0$.

We have thus shown that $(\mathbb{R}^k, \mathcal{L}, \lambda)$ satisfies the first 6 claims of the theorem. Any set contained in the σ-algebra $\mathcal{L}$ is called a Lebesgue measurable set, and the measure λ is called the Lebesgue measure.

Step 3: The Volume of k-cells under the Lebesgue Measure

Now we must show that the Lebesgue measure of a k-cell is precisely that of its volume. This allows us to interpret the Lebesgue measure of any Lebesgue measurable set as the volume of that set, by making use of the fact that every nonempty open set is the countable union of k-cells.

Choose any nonempty open k-cell

$$W = (a_1, b_1) \times \cdots \times (a_k, b_k)$$

in $\mathbb{R}^k$. For any $n \in N_+$, let A_n be defined as

$$A_n = \bigcup_{Q \in \Omega_n, \overline{Q} \subset W} Q.$$

Clearly, $\{A_n\}_{n \in N_+}$ is an increasing sequence of sets. Let $Q \in \Omega_n$ and $\overline{Q} \subset W$; then, for any $x \in Q$, letting $Q' \in \Omega_{n+1}$ be the box in which x is contained, $Q' \subset Q$ by lemma 4.5, since $Q \cap Q' \neq \emptyset$. Furthermore, $\overline{Q'} \subset \overline{Q} \subset W$, so that

$$x \in Q' \subset \bigcup_{Q'' \in \Omega_{n+1}, Q'' \subset W} Q'' = A_{n+1}.$$

This holds for any $x \in Q$, so $Q \subset A_{n+1}$, and by implication $A_n \subset A_{n+1}$.

Additionally, $\bigcup_n A_n = W$. Each E_n, being the union of sets that lie in W, is clearly contained in W, so that $\bigcup_n A_n \subset W$. To see the reverse inclusion, choose any $x \in W$. $\{x\}$ is a compact set contained in the open set W, and $(\mathbb{R}^k, \tau_k)$ is a locally compact Hausdorff space, so by theorem 1.14, there exists an open set $V \in \tau_k$ with compact closure such that

$$x \in V \subset \overline{V} \subset W.$$

By lemma 4.5, V, being a nonempty open subset of $\mathbb{R}^k$, can be expressed as the countable union of sets in $\bigcup_n \Omega_n$. Choosing any $N \in N_+$ and $Q \in \Omega_N$ such that $x \in Q \subset V$, we can see that

$$\overline{Q} \subset \overline{V} \subset W,$$

so that $x \in Q \subset A_N$. This holds for any $x \in W$, so $W \subset \bigcup_n A_n$, and as such $W = \bigcup_n A_n$.

We can also easily see that $\text{vol}(A_n) \nearrow \text{vol}(W)$. For any $n \in N_+$, because $A_n \subset A_{n+1}$, we have $\text{vol}(A_n) \leq \text{vol}(A_{n+1})$, and if

$$\text{vol}(W) - \text{vol}(A_n) > 2^{-kn}$$

for some $n \in N_+$, this means that there exists at least one box in Ω_n whose closure lies in W but which is not included in the union A_n, a contradiction. Therefore,

$$|\text{vol}(W) - \text{vol}(A_n)| = \text{vol}(W) - \text{vol}(A_n) \leq 2^{-kn}$$

for any $n \in N_+$, which implies that

$$\text{vol}(A_n) \nearrow \text{vol}(W).$$

For any $n \in N_+$, because $\overline{A_n}$ is a closed set such that

$$\overline{A_n} = \overline{\bigcup_{Q \in \Omega_n, \overline{Q} \subset W} Q} \subset \bigcup_{Q \in \Omega_n, Q \subset W} \overline{Q} \subset W,$$

$\overline{A_n}$ is compact (it is closed and bounded, since W is bounded), and by Urysohn's lemma there exists an $f_n \in C_c(\mathbb{R}^k)$ such that

$$\overline{A_n} \prec f_n \prec W.$$

Define $g_n = \max(f_1, \cdots, f_n)$ for any $n \in N_+$. For any $n \in N_+$, we saw from the analysis in step 1 that

$$\text{vol}(A_n) = \Lambda_n I_{A_n} = \Lambda_m I_{A_n}$$

for any $m \geq n$. Because $I_{A_n} \leq f_n \leq g_n \leq I_W$, from the monotonicity of the linear function Λ_m for any $m \geq n$ we can see that

$$\text{vol}(A_n) = \Lambda_m I_{A_n} \leq \Lambda_m f_n,$$

so taking $m \to \infty$ on both sides yields

$$\text{vol}(A_n) \leq \Lambda f_n.$$

Furthermore, because $g_n(x) = 0$ for any $x \notin W$ and $0 \leq g_n(x) \leq 1$ for any $x \in \mathbb{R}^k$, for any $m \in N_+$ we have

$$\Lambda_m g_n = 2^{-km} \cdot \sum_{x \in P_m} g_n(x) \leq 2^{-km} |W \cap P_m| \leq \prod_{i=1}^{k} (b_i - a_i + 2^{-m}),$$

where the last inequality follows because $2^{-km}|W \cap P_m|$ is the sum of the volume all boxes in Ω_n whose lower left vertex is contained in W. Therefore,

$$\Lambda g_n = \lim_{m\to\infty} \Lambda_m g_n \leq= \prod_{i=1}^{k}(b_i - a_i) = \text{vol}(W),$$

and as such, by the montonicity of Λ,

$$\text{vol}(A_n) \leq \Lambda f_n \leq \Lambda g_n \leq \text{vol}(W).$$

The fact that $\text{vol}(A_n) \nearrow \text{vol}(W)$ indicates that

$$\lim_{n\to\infty} \Lambda g_n = \text{vol}(W)$$

as well.

$\{g_n\}_{n\in N_+}$ is an increasing sequence of functions, and because $I_{A_n} \nearrow I_W$ and $I_{A_n} \leq g_n \leq I_W$, $g_n \nearrow I_W$ as well. By the MCT,

$$\begin{aligned}\lambda(W) = \int_{\mathbb{R}^k} I_W d\lambda &= \lim_{n\to\infty} \int_{\mathbb{R}^k} g_n d\lambda \\ &= \lim_{n\to\infty} \Lambda g_n = \text{vol}(W).\end{aligned}$$

Now let W' be the closed k-cell $W' = [a_1, b_1] \times \cdots \times [a_k, b_k]$. Then, W' is the intersection of the open k-cells $\prod_{i=1}^{k}\left(a_i - \frac{1}{n}, b_i + \frac{1}{n}\right)$ across $n \in N_+$, and because

$$\lambda\left(\prod_{i=1}^{k}(a_i - 1, b_i + 1)\right) = \prod_{i=1}^{k}(b_i - a_i + 2) < +\infty,$$

by the sequential continuity of measures we have

$$\lambda(W') = \lim_{n\to\infty} \lambda\left(\prod_{i=1}^{k}\left(a_i - \frac{1}{n}, b_i + \frac{1}{n}\right)\right) = \lim_{n\to\infty}\prod_{i=1}^{k}\left(b_i - a_i + \frac{2}{n}\right) = \prod_{i=1}^{k}(b_i - a_i) = \text{vol}(W').$$

Finally, for any arbitrary k-cell W'' with vertices $(a_1, \cdots, a_k)$ and $(b_1, \cdots, b_k)$, since

$$W \subset W'' \subset W',$$

the montonicity of measures tells us that

$$\text{vol}(W) = \lambda(W) \leq \lambda(W'') \leq \lambda(W') = \text{vol}(W').$$

Because $\text{vol}(W) = \text{vol}(W') = \prod_{i=1}^{k}(b_i - a_i)$, we have $\lambda(W'') = \prod_{i=1}^{k}(b_i - a_i)$, which is exactly the volume of W''.

Step 4: The Translation-Invariance of the Lebesgue Measure

We show this property by first starting with boxes and then gradually generalizing the result, first to open sets, then to Borel sets and finally to Lebesgue measurable sets.

Let Q be a half-open k-cell given by

$$Q = [a_1, b_1) \times \cdots \times [a_k, b_k).$$

Then, for any $x \in \mathbb{R}^k$,

$$Q + x = \{y + x \mid y \in Q\} = \prod_{i=1}^{k} [a_i + x_i, b_i + x_i).$$

Since Q and $Q + x$ are k-cells, from the above result we have

$$\lambda(Q + x) = \mathrm{vol}(Q + x) = \prod_{i=1}^{k} (b_i - a_i) = \mathrm{vol}(Q) = \lambda(Q).$$

Now let $V \in \tau_k$ be nonempty. We saw in lemma 4.5 that V can be expressed as the countable union of disjoint half open k-cells; that is, there exists a countable collection $\{W_n\}_{n \in N_+}$ of disjoint half-open k-cells such that

$$V = \bigcup_n W_n.$$

Then, for any $x \in \mathbb{R}^k$,

$$V + x = \bigcup_n (W_n + x),$$

and by the disjointness of $\{W_n\}_{n \in N_+}$ and countable additivity, we have

$$\lambda(V + x) = \sum_{n=1}^{\infty} \lambda(W_n + x) = \sum_{n=1}^{\infty} \lambda(W_n) = \lambda(V).$$

For any $A \in \mathcal{L}$ and $\varepsilon > 0$, there exist a closed set F and an open set V such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) < \varepsilon.$$

For any $x \in \mathbb{R}^k$, since $V \setminus F$ is an open set,

$$\lambda((V \setminus F) + x) = \lambda(V \setminus F) < \varepsilon,$$

and as such,

$$\begin{aligned} \lambda(A) \leq \lambda(V) = \lambda(V + x) &= \lambda((V \setminus F) + x) + \lambda(F + x) < \varepsilon + \lambda(A + x) \\ \lambda(A + x) \leq \lambda(V + x) = \lambda(V) &= \lambda(V \setminus F) + \lambda(F) \leq \varepsilon + \lambda(A), \end{aligned}$$

so that

$$|\lambda(A) - \lambda(A + x)| < \varepsilon.$$

This holds for any $\varepsilon > 0$, so

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

Step 5: Translation Invariant Borel Measures as Scaled Lebesgue Measures

Let μ be a measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ that is translation-invariant, that is,

$$\mu(A) = \mu(A + x)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$ and $x \in \mathbb{R}^k$.
Note that any two δ-boxes $Q(x;\delta)$ and $Q(y;\delta)$ are translations of one another. Specifically,

$$Q(x;\delta) = Q(y;\delta) + (x - y).$$

Therefore, the measure of any δ-box under a translation invariant measure must be the same.

Choose any 1-box Q_0, and define

$$c = \mu(Q_0) \geq 0,$$

where c does not depend on the vertices of Q_0 because μ is translation invariant. Note that $\lambda(Q_0) = 1$.
Let Q be an arbitrary 2^{-n}-box; since Q_0 is the union of 2^{nk} disjoint 2^{-n}-boxes, by finite additivity and the translation invariance of μ and λ we have

$$2^{nk}\mu(Q) = \mu(Q_0) = c \cdot \lambda(Q_0) = c \cdot 2^{nk}\lambda(Q),$$

which implies that

$$\mu(Q) = c \cdot \lambda(Q).$$

We have shown above that $\mu = c \cdot \lambda$ on $\bigcup_n \Omega_n$. Let $V \in \tau_k$ be nonempty. We saw in lemma 4.5 that V can be expressed as the countable union of disjoint k-cells in $\bigcup_n \Omega_n$; that is, there exists a countable collection $\{W_n\}_{n \in N_+}$ of disjoint k-cells in $\bigcup_n \Omega_n$ such that

$$V = \bigcup_n W_n.$$

By the disjointness of $\{W_n\}_{n\in N_+}$ and countable additivity, we have

$$\mu(V) = \sum_{n=1}^{\infty} \mu(W_n) = c\sum_{n=1}^{\infty} \lambda(W_n) = c\cdot\lambda(V).$$

For any $A \in \mathcal{B}(\mathbb{R}^k)$ and $\varepsilon > 0$, because $\mathcal{B}(\mathbb{R}^k) \subset \mathcal{L}$, there exist a closed set F and an open set V such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) < \varepsilon.$$

For any $x \in \mathbb{R}^k$, since $V \setminus F$ is an open set,

$$\begin{aligned} c\cdot\lambda(V\setminus F) = \mu(V\setminus F) &< c\cdot\varepsilon \\ c\cdot\lambda(V) &= \mu(V), \end{aligned}$$

and as such,

$$\begin{aligned} c\cdot\lambda(A) \le c\cdot\lambda(V) = \mu(V) = \mu(V\setminus F) + \mu(F) < c\cdot\varepsilon + \mu(A) \\ \mu(A) \le \mu(V) = c\cdot\lambda(V) = c\cdot\lambda(V\setminus F) + c\cdot\lambda(F) \le c\cdot\varepsilon + c\cdot\lambda(A), \end{aligned}$$

so that

$$|c\cdot\lambda(A) - \mu(A)| < c\cdot\varepsilon.$$

This holds for any $\varepsilon > 0$, so

$$\mu(A) = c\cdot\lambda(A).$$

Step 6: The Lebesgue Measure under Linear Transformations

To show the final property of the Lebesgue measure, let T be a $k \times k$ real matrix. Suppose initially that the range $R(T)$ of T is $\mathbb{R}^k$, so that T has full rank. This means that T is an invertible linear operator, and because the inverse of a linear operator is also linear, this makes T a homeomorphism on $\mathbb{R}^k$. Therefore, letting $U = T^{-1}$,

$$T(A) = U^{-1}(A) \in \mathcal{B}(\mathbb{R}^k)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$ by the definition of continuity, and we can define $\mu : \mathcal{B}(\mathbb{R}^k) \to [0, +\infty]$ as

$$\mu(A) = \lambda(T(A))$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$. μ is a Borel measure on $\mathbb{R}^k$:

- $\mu(\emptyset) = \lambda(\emptyset) = 0$, and
- For any disjoint $\{A_n\}_{n \in N_+} \subset \mathcal{B}(\mathbb{R}^k)$ with $A = \bigcup_n A_n$, we have

  $$T(A) = \bigcup_n T(A_n),$$

  where $\{T(A_n)\}_{n \in N_+}$ is a sequence of disjoint Borel sets, so

  $$\mu(A) = \lambda(T(A)) = \sum_{n=1}^{\infty} \lambda(T(A_n)) = \sum_{n=1}^{\infty} \mu(A_n)$$

  by the countable additivity of λ.

Furthermore, because of the translation invariance of λ,

$$\mu(A + x) = \lambda(T(A + x)) = \lambda(T(A) + Tx) = \lambda(T(A)) = \mu(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$ and $x \in \mathbb{R}^k$, so that μ is translation invariant as well.
Finally, for any compact $K \subset \mathbb{R}^k$, because $T(K)$ is also compact by the continuity of T, we can see that

$$\mu(K) = \lambda(T(K)) < +\infty.$$

μ is a Borel measure on $\mathbb{R}^k$ that is translation invariant and assigns finite measure to any compact set. By the preceding result, there exists a $c \in \mathbb{R}_+$ such that

$$\mu(A) = \lambda(T(A)) = c \cdot \lambda(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$. If $c = 0$, then

$$\mu(\mathbb{R}^k) = \lambda(\mathbb{R}^k) = 0,$$

a contradiction, so $c > 0$. Denoting $c = \Delta(T)$, it follows that

$$\lambda(T(A)) = \Delta(T) \cdot \lambda(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^k)$.

Choosing any $A \in \mathcal{L}$, there exist Borel sets F, V such that $F \subset A \subset V$ and

$$\lambda(V \setminus F) = 0.$$

Then, because $T(F) \subset T(A) \subset T(V)$,

$$\Delta(T) \cdot \lambda(F) \leq \lambda(T(A)) \leq \Delta(T) \cdot \lambda(V),$$

where

$$\lambda(V) = \lambda(V \setminus F) + \lambda(F) = \lambda(F),$$

which implies $\lambda(A) = \lambda(F) = \lambda(V)$.
Therefore,

$$\lambda(T(A)) = \Delta(T) \cdot \lambda(V) = \Delta(T) \cdot \lambda(A).$$

Finally, suppose that the range of T is a proper subspace of $\mathbb{R}^k$. Let $\{v_1, \cdots, v_m\}$ be an orthonormal basis of $T(\mathbb{R}^k)$, where $m < k$, and $\{v_1, \cdots, v_k\}$ the extension of $\{v_1, \cdots, v_m\}$ to an orthonormal basis of $\mathbb{R}^k$. Then, defining S as the unique linear operator on $\mathbb{R}^k$ such that $Se_i = v_i$ for $1 \leq i \leq k$, where $\{e_1, \cdots, e_k\}$ is the standard basis of $\mathbb{R}^k$, we can see that

$$T(\mathbb{R}^k) = S\left(\mathbb{R}^m \times \{\mathbf{0}_{(k-m)\times 1}\}\right).$$

Therefore,

$$\lambda(T(\mathbb{R}^k)) = \lambda(S\left(\mathbb{R}^m \times \{\mathbf{0}_{(k-m)\times 1}\}\right)) = \Delta(S) \cdot \lambda(\mathbb{R}^m \times \{\mathbf{0}_{(k-m)\times 1}\}) = 0.$$

This means that

$$\lambda(T(A)) = 0$$

for any $A \in \mathcal{L}$, and as such that $\Delta(T) = 0$.

Q.E.D.

It follows directly from the construction that the Lebesgue and Riemann integrals of continuous compactly supported funtions must be the same.

4.3.1 The Determinant of Square Matrices

The value $\Delta(T)$ above has an interpretation as the determinant of the matrix T.

For any matrix $A \in \mathbb{R}^{k\times k}$, denoting the (k,j)th element of A by $A(k,j)$, we define the determinant of A as

$$\det(A) = \sum_{j=1}^{n}(-1)^{1+j}A(1,j)\cdot\det\left(\tilde{A}_{1j}\right),$$

where $\tilde{A}_{ij}$ is the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and jth column of A. Here, we implicitly put the determinant of a dimension-less matrix equal to 1, which implies that the determinant of a scalar is just itself. We can consider the determinant a function of the n columns $A_1,\cdots,A_n$ of A, and write $\det(A) = \det(A_1,\cdots,A_n)$. The determinant has the following properties:

Lemma 4.8 (Properties of the Determinant)
Let $A \in \mathbb{R}^{n\times n}$. Then, the following hold true:

i) **(n-linearity)** $\det(A)$ is linear with respect to each column of A.

ii) **(Alternating Property)** If B is obtained by interchanging two columns of A, then $\det(B) = -\det(A)$. By implication, if A has two identical columns, then $\det(A) = 0$.

iii) **(Assigns Value 1 to the Identity Matrix)** $\det(I_n) = 1$.

iv) If A is singular, then $\det(A) = 0$.

Proof) i) We proceed by induction. The statement is obviously true when $n = 1$, so suppose it holds for some $n \geq 1$. Choose any $v_1,\cdots,v_{n+1},u \in \mathbb{R}^{n+1}$ and $a \in \mathbb{R}$, and denote by $\tilde{v}_k$, $\tilde{u}$ the vectors v_k and u obtained by deleting their first element. Fix $1 \leq j \leq n$. For any $1 \leq k \leq n+1$, denote by $\tilde{V}_{-k}$ the $n\times n$ matrix with columns equal to $\tilde{v}_1,\cdots,\tilde{v}_{k-1},\tilde{v}_{k+1},\cdots,\tilde{v}_{n+1}$.
Similarly, if $k \neq j$, let $\tilde{V}_{-k,u}$ be the $n\times n$ matrix obtained by replacing $\tilde{v}_j$ with $\tilde{u}$ in $\tilde{V}_{-k}$, and let $\tilde{V}_{-k,av+u}$ be the $n\times n$ matrix obtained by replacing $\tilde{v}_j$ with $a\tilde{v}_j + \tilde{u}$ in $\tilde{V}_{-k}$.

Then,

$$\begin{aligned}\det(v_1,\cdots,av_j+u,\cdots,v_{n+1}) = &\sum_{k\neq j}(-1)^{1+k}v_k(1)\cdot\det\left(\tilde{V}_{-k,av+u}\right)\\ &+(-1)^{i+j}(a\cdot v_j(i)+u(i))\cdot\det\left(\tilde{V}_{-j}\right).\end{aligned}$$

By the inductive hypothesis, for any $1 \leq k \leq n+1$ such that $k \neq j$,

$$\det\left(\tilde{V}_{-k,av+u}\right) = a\cdot\det\left(\tilde{V}_{-k}\right) + \det\left(\tilde{V}_{-k,u}\right),$$

so we have

$$
\begin{aligned}
\det(v_1,\cdots,av_j+u,\cdots,v_{n+1}) &= a\cdot\sum_{k\neq j}(-1)^{1+k}v_k(1)\cdot\det\left(\tilde{V}_{-k}\right)+\sum_{k\neq j}(-1)^{1+k}v_k(1)\cdot\det\left(\tilde{V}_{-k,n}\right)\\
&\quad +a\cdot(-1)^{1+j}v_j(1)\cdot\det\left(\tilde{V}_{-j}\right)+(-1)^{1+j}u(1)\cdot\det\left(\tilde{V}_{-j}\right)\\
&= a\cdot\sum_{k=1}^{n+1}(-1)^{1+k}v_k(1)\cdot\det\left(\tilde{V}_{-k}\right)\\
&\quad +\left[\sum_{k\neq j}(-1)^{1+k}v_k(1)\cdot\det\left(\tilde{V}_{-k,n}\right)+(-1)^{1+j}u(1)\cdot\det\left(\tilde{V}_{-j}\right)\right]\\
&= a\cdot\det(v_1,\cdots,v_{n+1})+\det(v_1,\cdots,u,\cdots,v_{n+1}).
\end{aligned}
$$

It now follows from induction that the determinant of $A\in\mathbb{R}^{n\times n}$ is linear in each of the columns of A for any $n\in N_+$.

ii) We initially proceed by induction. The claim trivially holds true for $n=1$, since in that case there is only one column and thus nothing to interchange it with. Now suppose the claim holds for some $n\geq 1$. Let $A=(A_1,\cdots,A_{n+1})\in\mathbb{R}^{(n+1)\times(n+1)}$, and construct B by interchaging the jth and $j+1$th rows of A, where $1\leq j<n+1$, that is,

$$B=(A_1,\cdots,A_{j+1},A_j,\cdots,A_{n+1}).$$

Then,

$$
\begin{aligned}
\det(B) = &\sum_{i\neq j,j+1}(-1)^{1+i}A(1,i)\cdot\det\left(\tilde{B}_{1i}\right)\\
&+(-1)^{1+j}A(1,j+1)\cdot\det\left(\tilde{A}_{1,j+1}\right)+(-1)^{1+j+1}A(1,j)\cdot\det\left(\tilde{A}_{1j}\right).
\end{aligned}
$$

For any $i\neq j,j+1$, $\tilde{B}_{1i}$ is an $n\times n$ matrix obtained by interchanging the jth and $j+1$th columns of $\tilde{A}_{1i}$, so by the inductive hypothesis,

$$\det\left(\tilde{B}_{1i}\right)=-\det\left(\tilde{A}_{1i}\right).$$

Therefore,

$$\det(B)=-\sum_{i=1}^{n+1}(-1)^{1+i}A(1,i)\cdot\det\left(\tilde{A}_{1i}\right)=-\det(A).$$

Now let B be obtained from A by interchanging any two distinct columns, say, the jth and kth columns, where we assume $j<k$. Then, B can be obtained from A in

an odd number of steps by interchanging adjacent columns. Specifically, we obtain B_1 by interchanging the jth and $j+1$th columns of A, then B_2 by interchanging the $j+1$th and $j+2$th columns of B_2, and so on and so forth, until we obtain B_{k-j} by interchanging the $k-1$th and kth columns of B_{k-j-1}. This sends column j to the kth position. To move column k to the jth position, we move backward by interchanging the $k-1$th and $k-2$th columns of B_{k-j} to obtain B_{k-j+1}, and so on and so forth until we interchange the $j+1$th and jth columns of $B_{2(k-j)-2}$ to obtain $B = B_{2(k-j)-1}$. The subscripts represent how many times the adjacent columns of A must be interchanged to obtain the matrix, and thus we can see that it requires us to interchange $2(k-j)-1$ adjacent columns of A to obtain B. Therefore, by the result we just proved on interchanging adjacent columns,

$$\det(B) = (-1)^{2(k-j)-1} \cdot \det(A) = -\det(A).$$

By induction, the claim holds for matrices of any dimension $n \in N_+$.

Suppose now that A has two identical columns. Then, the determinant of the matrix $\tilde{A}$ obtained by interchanging those columns is equal to $-\det(A)$. However, since $\tilde{A} = A$, we have

$$\det\left(\tilde{A}\right) = \det(A) = -\det(A).$$

This implies that $2 \cdot \det(A) = 0$ and thus $\det(A) = 0$.

iii) This can also be seen by induction. It obviously holds true for $n = 1$, so suppose it holds for some $n \geq 1$. Then, denoting $A = I_{n+1}$,

$$\det(A) = A(1,1) \cdot \det\left(\tilde{A}_{11}\right).$$

Here, $A(1,1) = 1$ and $\tilde{A}_{11}$ is the $n \times n$ identity matrix, so we can conclude that

$$\det(A) = \det(I_{n+1}) = 1.$$

iv) The last claim is an application of the n-linearity and alternating property of the determinant. It is trivial for $n = 1$, so let $n > 1$. Suppose that A is a singular $n \times n$ matrix. Let $A_1, \cdots, A_n \in \mathbb{R}^n$ be the columns of A; by definition, there exists a non-zero vector $x \in \mathbb{R}^n$ such that $Ax = \mathbf{0}$. Assuming without loss of generality that $x_1 \neq 0$, we can see that

$$A_1 = \sum_{i=2}^{n} \left(-\frac{x_i}{x_1}\right) A_i,$$

so that A_1 can be written as the linear combination of the other $n-1$ columns. By the n-linearity of the determinant, we have

$$\det(A) = \sum_{i=2}^{n} \left(-\frac{x_i}{x_1}\right) \cdot \det(A_i, A_2, \cdots, A_n).$$

Since each $\det(A_i, A_2, \cdots, A_n)$ is the determinant of a matrix with identical columns, by the alternating property we can conclude that $\det(A) = 0$.

Q.E.D.

Of special interest is the determinant of elementary matrices, that is, matrices that are obtained either by:

Type 1 : Interchanging the columns of I_n

Type 2 : Multiplying the column of I_n by a constant $c \neq 0$

Type 3 : Adding a column of I_n to another column of I_n

The determinant of these matrices can be easily obtained by using the n-linearity and alternating properties of the determinant, alongside the fact that $D(I_n) = 1$. Let $\{e_1, \cdots, e_n\} \subset \mathbb{R}^n$ be the standard basis of $\mathbb{R}^n$, and consider an elementary matrix E. If E is of type 1, then it is obtained by interchanging two columns of I_n; therefore, by the alternating property,

$$\det(E) = -\det(I_n) = -1.$$

If E is of type 2, then there exists a $1 \leq j \leq n$ and $c \neq 0$ such that

$$E = \begin{pmatrix} e_1 & \cdots & c \cdot e_j & \cdots & e_n \end{pmatrix},$$

so by the n-linearity of the determinant,

$$\det(E) = c \cdot \det(I_n) = c.$$

Finally, if E is of type 3, then there exist $1 \leq j \neq k \leq n$ such that

$$E = \begin{pmatrix} e_1 & \cdots & \underbrace{e_j + e_k}_{j\textbf{th position}} & \cdots e_n \end{pmatrix},$$

so by the n-linearity of the determinant and the alternating property,

$$\det(E) = \det(I_n) + \det(e_1, \cdots, e_k, \cdots, e_k, \cdots, e_n) = \det(I_n) = 1.$$

In any case, $\det(E) \neq 0$.

It can actually be shown that the determinant is the unique function with the first three properties stated above. This is shown formally in the theorem below, and in the process, we manage to obtain an alternate characterization of determinants.

Theorem 4.9 (Uniqueness of the Determinant)
Let $D : \mathbb{R}^{n\times n} \to \mathbb{R}$ be a function that is n-linear with respect to each column and has the alternating property. Then, for any $A \in \mathbb{R}^{n\times n}$,

$$D(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{n} A(\sigma i, i) \right) D(I_n),$$

where the sum runs over the set S_n of all permutations of $\{1, \cdots, n\}$, and we define

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is obtained from } \{1, \cdots, k\} \text{ via an even number of transpositions} \\ 0 & \text{otherwise} \end{cases}$$

for any $\sigma \in S_n$.

Proof) Let $\{e_1, \cdots, e_n\} \subset \mathbb{R}^n$ be the standard basis of $\mathbb{R}^n$, and choose any $A \in \mathbb{R}^{n\times n}$. Denoting the columns of A by $A_1, \cdots, A_n \in \mathbb{R}^n$, note that

$$A_j = \sum_{i=1}^{n} A(i, j) e_i$$

for any $1 \leq j \leq n$. Then, by the n-linearity of D,

$$\begin{aligned} D(A) = D(A_1, \cdots, A_n) &= D\left(\sum_{i_1=1}^{n} A(i_1, 1) e_{i_1}, A_2, \cdots, A_n \right) \\ &= \sum_{i_1=1}^{n} A(i_1, 1) \cdot D(e_{i_1}, A_2, \cdots, A_n). \end{aligned}$$

Likewise,

$$\begin{aligned} D(e_{i_1}, A_2, \cdots, A_n) &= D\left(e_{i_1}, \sum_{i_2=1}^{n} A(i_2, 2) e_{i_2}, \cdots, A_n \right) \\ &= \sum_{i_2=1}^{n} A(i_2, 2) \cdot D(e_{i_1}, e_{i_2}, \cdots, A_n) \end{aligned}$$

for any $1 \le i_1 \le n$, and continuing on in this manner leads us to

$$D(A) = \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} A(i_1,1)\cdots A(i_n,n) D(e_{i_1},\cdots,e_{i_n}).$$

Since any matrix with identical rows has value 0 under D due to the alternating property,

$$D(e_{i_1},\cdots,e_{i_n}) = 0$$

if $i_j = i_k$ for any two $1 \le j \ne k \le n$. Therefore,

$$D(A) = \sum_{\sigma \in S_n} A(\sigma 1,1)\cdots A(\sigma n,n) D(e_{\sigma 1},\cdots,e_{\sigma n}).$$

For any $\sigma \in S_n$, if $\operatorname{sgn}(\sigma) = 1$, then the matrix

$$\begin{pmatrix} e_{\sigma 1} & \cdots & e_{\sigma n} \end{pmatrix}$$

can be obtained from interchanging the columns of I_n an even number of times. Thus, by the alternating property, we can see that

$$D(e_{\sigma 1},\cdots,e_{\sigma n}) = D(I_n).$$

On the other hand, if $\operatorname{sgn}(\sigma) = -1$, then

$$D(e_{\sigma 1},\cdots,e_{\sigma n}) = -D(I_n).$$

It follows that

$$D(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)\left(A(\sigma 1,1)\cdots A(\sigma n,n)\right) D(I_n),$$

which is exactly the desired result.

Q.E.D.

Using the general formula for n-linear and alternating functions on $\mathbb{R}^{n\times n}$, we can now furnish an alternate formula for determinants. This formulation can be used to easily derive two core properties of determinants.

Theorem 4.10 (Leibniz Formula for Determinants)
For any $A \in \mathbb{R}^{n\times n}$,

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\left(\prod_{i=1}^{n} A(\sigma i, i)\right).$$

Furthermore, the determinant also possesses the following properties:

i) **(Invariance under Transposition)** For any $A \in \mathbb{R}^{n\times n}$, $\det(A) = \det(A')$.

ii) **(Separability under Matrix Products)** For any $A, B \in \mathbb{R}^{n\times n}$, $\det(AB) = \det(A)\det(B)$.

Proof) We showed in lemma 4.8 that the determinant is n-linear with respect to each column, has the alternating property, and assigns value 1 to I_n. Therefore, by theorem 4.9,

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\left(\prod_{i=1}^{n} A(\sigma i, i)\right)$$

for any $A \in \mathbb{R}^{n\times n}$.

To show the second result, we must introduce the concept of inverse permutations. Given any $\sigma \in S_n$, we can also define $\sigma^{-1}j$ for any $1 \le j \le n$; $\sigma^{-1}j$ is the position of the number j under the permutation σ, while σj is the number assigned to position j under the permutation σ. Clearly, $\sigma\sigma^{-1}j = j$ for any $1 \le j \le n$.

Now choose any $A \in \mathbb{R}^{n\times n}$. By the Leibniz formula,

$$\begin{aligned}\det(A') &= \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\left(\prod_{i=1}^{n} A_{i,\sigma i}\right)\\ &= \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\left(\prod_{i=1}^{n} A_{\sigma^{-1}i,i}\right).\end{aligned}$$

Since the collection of all inverse permutations σ^{-1} of permutations in S_n is exactly the set S_n of all permutations of $\{1, \cdots, n\}$, it follows that

$$\det(A') = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)\left(\prod_{i=1}^{n} A_{\sigma i,i}\right) = \det(A).$$

Let $A \in \mathbb{R}^{n\times n}$, and define the function $D : \mathbb{R}^{n\times n} \to \mathbb{R}$ as

$$D(B) = \det(AB)$$

for any $B \in \mathbb{R}^{n\times n}$. D is n-linear with respect to each column and has the alternating property; to see n-linearity, note that, for any $a \in \mathbb{R}$, $v_1, \cdots, v_n, u \in \mathbb{R}^n$ and $1 \leq j \leq n$,

$$\begin{aligned} D(v_1, \cdots, a \cdot v_j + u, \cdots, v_n) &= \det\left(A\begin{pmatrix} v_1 & \cdots & a \cdot v_j + u & \cdots v_n \end{pmatrix}\right) \\ &= \det(Av_1, \cdots, a \cdot (Av_j) + Au, \cdots, Av_n) \\ &= a \cdot \det(Av_1, \cdots, Av_n) + \det(Av_1, \cdots, Au, \cdots, Av_n) \\ &= a \cdot \det\left(A\begin{pmatrix} v_1 & \cdots v_n \end{pmatrix}\right) + \det\left(A\begin{pmatrix} v_1 & \cdots & u & \cdots v_n \end{pmatrix}\right) \\ &= a \cdot D(v_1, \cdots, v_n) + D(v_1, \cdots, u, \cdots, v_n), \end{aligned}$$

where the third equality follows from the n-linearity of the determinant. For the alternating property, note that, for any $B \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{n \times n}$ obtained by interchanging columns j and k of B, the matrix AC is obtained by also interchanging columns j and k of AB. From the alternating property of the determinant, we now have

$$D(C) = \det(AC) = -\det(AB) = D(B).$$

Since D is n-linear with respect to each column and has the alternating property, by the theorem above

$$D(B) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{n} B(\sigma i, i) \right) D(I_n) = \det(B) D(I_n)$$

for any $B \in \mathbb{R}^{n \times n}$. Since $D(I_n) = \det(A)$ in this case, we can see that, by the Leibniz formula for determinants,

$$\det(AB) = D(B) = \det(B) D(I_n) = \det(B) \det(A).$$

Q.E.D.

Corollary to Theorem 4.10 (Characterization of Invertible Matrices)
Let $A \in \mathbb{R}^{n \times n}$. A is non-singular if and only if $\det(A) \neq 0$.

Proof) Sufficiency was shown in lemma 4.8; if A is singular then $\det(A) = 0$, so the contraposition must also be true, namely it must be the case that if $\det(A) \neq 0$, then A is non-singular.

For sufficiency, suppose that A is non-singular. Then, it can be expressed as a finite product of elementary matrices, that is,

$$A = \prod_{i=1}^{k} E_i$$

where each $E_i \in \mathbb{R}^{n\times n}$ is an elementary matrix. We saw previously that the determinant of an elementary matrix is always non-zero; therefore, using the product rule for determinants,

$$\det(A) = \prod_{i=1}^{k} \det(E_i) \neq 0.$$

This completes the proof.

Q.E.D.

That the determinant of a matrix and its transpose is the same also has many important implications. For one, it shows that the alternating property holds even when the rows of a matrix are interchanged; to see this, let $A \in \mathbb{R}^{n\times n}$, and suppose B is obtained by interchanging two rows of A. Then, B' is obtained from A' by interchanging two columns, so

$$\det(B) = \det(B') = -\det(A') = -\det(A)$$

by applying the alternating property and the above result. From this it follows that any matrix with identical rows has determinant 0.

Perhaps the most useful part of the above result is that it helps expand our initial definition of the determinant. While our initial definition of $\det(A)$ is stated in terms of an expansion with respect to the first row, we can also furnish an equivalent expansion with respect to the first column. In fact, this expansion can be along any row or column, which we prove below:

Lemma 4.11 Let $A \in \mathbb{R}^{n\times n}$. Then,

$$\det(A) = \sum_{k=1}^{n} (-1)^{i+k} A(i,k) \cdot \det\left(\tilde{A}_{ik}\right) = \sum_{k=1}^{n} (-1)^{k+j} A(k,j) \cdot \det\left(\tilde{A}_{kj}\right)$$

for any $1 \leq i \leq n$ and $1 \leq j \leq n$.

Proof) Since the result follows by definition if $i = 1$, fix $1 < i \leq n$. Let B be the $n \times n$ matrix constructed by moving the ith row up to the first row and pushing the first to the $i-1$th row down by one row. Because we obtained B by interchanging the rows of A $i-1$ times, we have

$$\det(B) = (-1)^{i-1} \det(A).$$

Since $\tilde{B}_{1j} = \tilde{A}_{ij}$ for any $1 \le j \le n$, by definition we have

$$\det(A) = (-1)^{i-1} \cdot \det(B) = (-1)^{i-1} \cdot \sum_{j=1}^{n} (-1)^{1+j} B(1,j) \cdot \det\left(\tilde{B}_{1j}\right)$$
$$= \cdot \sum_{j=1}^{n} (-1)^{i+j} A(i,j) \cdot \det\left(\tilde{A}_{ij}\right).$$

For the column expansion, we need only use the fact that determinants are transposition invariant: for any $A \in \mathbb{R}^{n\times n}$ and $1 \le j \le n$,

$$\det(A) = \det(A') = \sum_{k=1}^{n} A'(j,k) \cdot \det\left(\tilde{A'}_{jk}\right)$$
$$= \sum_{k=1}^{n} A(k,j) \cdot \det\left(\left(\tilde{A}_{kj}\right)'\right) = \sum_{k=1}^{n} A(k,j) \cdot \det\left(\tilde{A}_{kj}\right).$$

Q.E.D.

The above result is more significant than it first seems, because it allows us to obtain a closed form expression for the inverse of a non-singular matrix.

Theorem 4.12 (Inverse of a Non-singular Matrix)
Let $A \in \mathbb{R}^{n\times n}$ be a non-singular matrix. Then, its inverse A^{-1} is given as

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A),$$

where $\operatorname{adj}(A)$, called the adjunct matrix of A, is an $n \times n$ matrix whose (i,j)th element is

$$(-1)^{i+j} \det\left(\tilde{A}_{ji}\right).$$

Proof) Define

$$B = \frac{1}{\det(A)} \operatorname{adj}(A),$$

where the adjunct of A is defined as above. We want to show that $BA = AB = I_n$. This follows from a simple computation. The (i,j)th element of AB is defined as

$$(AB)(i,j) = \sum_{k=1}^{n} A(i,k)B(k,j) = \frac{1}{\det(A)} \sum_{k=1}^{n} (-1)^{k+j} A(i,k) \cdot \det\left(\tilde{A}_{jk}\right).$$

If $i = j$, then

$$(AB)(i,j) = \frac{1}{\det(A)} \sum_{k=1}^{n} (-1)^{i+k} A(i,k) \cdot \det\left(\tilde{A}_{ik}\right) = \frac{\det(A)}{\det(A)} = 1$$

by a determinant expansion with respect to the ith row. On the other hand, if $i \neq j$, then

$$\sum_{k=1}^{n}(-1)^{k+j}A(i,k)\cdot\det\left(\tilde{A}_{jk}\right) = (-1)^{j-i}\cdot\sum_{k=1}^{n}(-1)^{k+i}A(i,k)\cdot\det\left(\tilde{A}_{jk}\right)$$

is the determinant of a matrix in which the ith row of A appears twice. Therefore, by the alternating property, this determinant equals 0, and we have

$$AB = I_n.$$

$BA = I_n$ follows from roughly the same steps, utilizing a column expansion rather than a row expansion.

Q.E.D.

4.3.2 The Geometric Meaning of the Determinant

Returning to the context of the Lebesgue measure, the determinant of a matrix has the geometric property of rescaling a set. Letting $\mathcal{L}$ be the set of all Lebesgue-measurable sets on $\mathbb{R}^k$, and λ the corresponding Lebesgue measure, recall that for any $T \in \mathbb{R}^{k\times k}$, there exists a $\Delta(T) \geq 0$ such that

$$\lambda(TA) = \Delta(T) \cdot \lambda(A)$$

for any $A \in \mathcal{L}$, where $\Delta(T) = 0$ if T is singular. To study the properties of the function $\Delta : \mathbb{R}^{k\times k} \to \mathbb{R}_+$, we fix $A = [0,1]^k$, which has volume 1 and thus measure 1 under the Lebesgue measure. This shows us that $\Delta(T)$ is given as

$$\Delta(T) = \lambda(T[0,1]^k).$$

Note that the operation Δ satisfies the same product rule as the determinant; that is, for any $U, T \in \mathbb{R}^{k\times k}$,

$$\begin{aligned}\Delta(UT) = \lambda((UT)[0,1]^k) = \lambda(U(T[0,1]^k)) &= \Delta(U) \cdot \lambda(T[0,1]^k) \\ &= \Delta(U)\Delta(T)\lambda([0,1]^k) = \Delta(U)\Delta(T).\end{aligned}$$

It turns out that $\Delta(T)$ shares more than just the product rule of the determinant. In fact, it is exactly equal to the absolute value of the determinant of T:

Theorem 4.13 (Geometric Meaning of the Determinant)
Let $\mathcal{L}$ be the set of all Lebesgue-measurable sets on $\mathbb{R}^k$, and λ the corresponding Lebesgue measure. Let $T \in \mathbb{R}^{k\times k}$. Then,

$$\lambda(T(A)) = |\det(T)| \cdot \lambda(A)$$

for any $A \in \mathcal{L}$.

Proof) In terms of the notations introduced above, we need only prove that

$$\Delta(T) = |\det(T)|$$

for any $T \in \mathbb{R}^{k\times k}$.

If T is singular, then $\Delta(T) = \det(T) = 0$, and the equivalence is established, so suppose T is non-singular. In this case, T can be expressed as the product of a finite number of elementary matrices, that is,

$$T = \prod_{i=1}^{n} E_i,$$

where each $E_i \in \mathbb{R}^{k\times k}$ is an elementary matrix. Thus, if we show that $\Delta(E) = |\det(E)|$ for any elementary matrix $E \in \mathbb{R}^{k\times k}$, then by the product rule,

$$|\det(T)| = \prod_{i=1}^{n} |\det(E_i)| = \prod_{i=1}^{n} \Delta(E_i) = \Delta(T),$$

and we have the equivalence. Let $E \in \mathbb{R}^{k\times k}$ be an elementary matrix, and $\{e_1, \cdots, e_k\} \subset \mathbb{R}^k$ the standard basis of $\mathbb{R}^k$. We consider the following three cases:

- **E is of Type 1**

  In this case, E is obtained by interchanging two columns of I_k, say, the ith and jth columns. Letting $i < j$ without loss of generality, this means that

  $$E = \begin{pmatrix} e_1 & \cdots & e_j & \cdots & e_i & \cdots e_k \end{pmatrix}.$$

  Note that $E[0,1]^k = [0,1]^k$. For any $x \in E[0,1]^k$, $Ey = x$ for some $y \in [0,1]^k$, where x is obtained by interchanging the ith and jth coordinates of y. Since the coordinates of $y \in [0,1]^k$ are all contained in $[0,1]$, so are the coordinates of x, and thus $x \in [0,1]^k$. To see the reverse inclusion, choose any $x \in [0,1]^k$, and, letting $y \in [0,1]^k$ be constructed by interchanging the ith and jth coordinates of x, $x = Ey$, so that $x \in E[0,1]^k$. It follows that

  $$\Delta(E) = \lambda(E[0,1]^k) = \lambda([0,1]^k) = 1 = |\det(E)|.$$

- **E is of Type 2**

  In this case, there exists a $c \neq 0$ and $1 \leq j \leq k$ such that

  $$E = \begin{pmatrix} e_1 & \cdots & c\cdot e_j & \cdots e_k \end{pmatrix}.$$

  Assume first that $c > 0$. We can show that $E[0,1]^k = \prod_{i=1}^{k}[0,c_i]$, where $c_i = 1$ for $i \neq j$ and $c_i = c$ for $i = j$. First, choose some $x \in E[0,1]^k$. Then, there exists a $y \in [0,1]^k$ such that $Ey = x$, and x is constructed by multiplying c to the jth element of y. Therefore, for any $1 \leq i \leq k$, $y_i = x_i \in [0,1]$, while $x_j = c \cdot y_j \in [0,c]$, so that

  $$x \in [0,1] \times \cdots \times [0,c] \times \cdots [0,1] = \prod_{i=1}^{k}[0,c_i].$$

Conversely, if $x \in \prod_{i=1}^{k}[0,c_i]$, then letting $y \in \mathbb{R}^k$ be defined as

$$y_i = \begin{cases} x_i & \text{if } i \neq j \\ \frac{1}{c}x_i & \text{if } i = j \end{cases},$$

we have $x = Ey \in E[0,1]^k$. It follows that

$$\Delta(E) = \lambda(E[0,1]^k) = \lambda\left(\prod_{i=1}^{k}[0,c_i]\right) = c = \det(E).$$

If $c < 0$, the same line of reasoning shows that

$$E[0,1]^k = [0,1] \times \cdots \times [c,0] \times \cdots [0,1],$$

so that $\Delta(E) = -c = |\det(E)|$.

– **E is of Type 3**

In this case, there exist $1 \leq i \neq j \leq k$ such that

$$E = \begin{pmatrix} e_1 & \cdots & \underbrace{e_i + e_j}_{j\text{th position}} & \cdots & e_k \end{pmatrix}.$$

For the sake of convenience, we work with the half-open unit cube $[0,1)^k$ instead of the unit cube $[0,1]^k$; note that they have the same volume 1.

For any $y \in \mathbb{R}^k$, denote $x = Ey$; then,

$$x_m = \begin{cases} y_m & \text{if } m \neq j \\ y_i + y_j & \text{if } m = j \end{cases}.$$

This means that, for any $y \in [0,1)^k$,

$$0 \leq x_m = y_m < 1 \quad \text{for any } m \neq j,$$
$$0 \leq x_j = y_i + y_j < y_i + 1 < 2.$$

We now partition $E[0,1)^k$ into two parts; defining

$$S_1 = \{x \in E[0,1)^k \mid x_j \in [0,1)\}, \quad S_2 = \{x \in E[0,1)^k \mid x_j \in [1,2)\},$$

$E[0,1]^k = S_1 \cup S_2$ and $S_1 \cap S_2 = \emptyset$. We can now see that

$$[0,1)^k = S_1 \cup (S_2 - e_j).$$

To see this, suppose $x \in [0,1)^k$. If $x_i \leq x_j$, then $x = Ey$ for $y \in [0,1)^k$ defined as

$$y_m = \begin{cases} x_m & \text{if } m \neq j \\ x_j - x_i & \text{if } m = j \end{cases}.$$

Thus, $x \in E[0,1)^k$, and since $x \in [0,1)^k$, we have $x \in S_1$. On the other hand, if $x_i > x_j$, then we can use the fact that the difference $x_j - x_i$ is bounded below by 1. Therefore, $x + e_j = Ey$ for $y \in [0,1)^k$ defined as

$$y_m = \begin{cases} x_m & \text{if } m \neq j \\ (x_j + 1) - x_i & \text{if } m = j \end{cases}.$$

Thus, $x + e_j \in E[0,1)^k$ and the jth element of $x + e_j$ is contained in the interval $[1,2)$, so we have $x + e_j \in S_2$ and $x \in S_2 - e_j$. This shows us that $[0,1)^k \subset S_1 \cup (S_2 - e_j)$.

Conversely, choose any $x \in S_1 \cup (S_2 - e_j)$. If $x \in S_1$, then $x \in [0,1)^k$ by definition. If $x \in (S_2 - e_j)$, then there exists a $y \in S_2$ such that $x = y - e_j$. By definition, $y_m \in [0,1)$ for any $m \neq j$ and $y_j \in [1,2)$, so $x \in [0,1)^k$, which shows us that $S_1 \cup (S_2 - e_j) \subset [0,1)^k$.

Finally, we can see that $S_1 \cap (S_2 - e_j) = \emptyset$. For the sake of contradiction, suppose that $x \in S_1 \cap (S_2 - e_j)$. Then, there exists a $w \in [0,1)^k$ such that $x = Ew - e_j$, since $x \in S_2 - e_j$. However, since $x \in S_1$, there also exists a $y \in [0,1)^k$ such that $x = Ey$. Putting these results together, we have

$$e_j = Ew - Ey = E(w - y).$$

Since the ith coordinate of $w - y$ is equal to 0, the above equation can be re-expressed as $e_j = w - y$, or $w_j = y_j + 1$. However, this implies that

$$w_j = y_j + 1 \geq 1,$$

which contradicts the fact that $w \in [0,1)^k$. Thus, $S_1 \cap (S_2 - e_j) = \emptyset$.

Now the result follows easily by noting that

$$\begin{aligned} 1 = \lambda([0,1)^k) &= \lambda(S_1 \cup (S_2 - e_j)) \\ &= \lambda(S_1) + \lambda(S_2 - e_j) && \text{(Finite Additivity)} \\ &= \lambda(S_1) + \lambda(S_2) && \text{(Translation Invariance of } \lambda\text{)} \\ &= \lambda(S_1 \cup S_2) && \text{(Finite Additivity)} \\ &= \lambda(E[0,1)^k) = \Delta(E). \end{aligned}$$

Since $\det(E) = 1$ as well, we have $\Delta(E) = |\det(E)|$ at last.

Q.E.D.

The above result can easily be extended to give the formula for integration under a linear change of variables:

Theorem 4.13 (Linear Change of Variables)
Let $\mathcal{L}$ be the set of all Lebesgue-measurable sets on $\mathbb{R}^k$, and λ the corresponding Lebesgue measure. Let $T \in \mathbb{R}^{k\times k}$ be a non-singular matrix and f a non-negative Lebesgue measurable function. Then,

$$\int_{\mathbb{R}^k} f(x)dx = |\det(T)| \cdot \int_{\mathbb{R}^k} f(T(x))dx.$$

Proof) We use the usual construction of non-negative functions using simple functions and indicator functions. Let f be a non-negative simple Lebesgue measurable function such that

$$f = \sum_{i=1}^{n} a_i \cdot I_{A_i},$$

where $a_1, \cdots, a_n \in [0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{L}$ are disjoint Lebesgue measurable subsets. Note that

$$f(T(x)) = \sum_{i=1}^{n} a_i \cdot I_{A_i}(T(x)) = \sum_{i=1}^{n} a_i \cdot I_{T^{-1}(A_i)}(x)$$

for any $x \in \mathbb{R}^k$, so that $f \circ T$ is also a non-negative simple Lebesgue measurable function. Then,

$$\begin{aligned}\int_{\mathbb{R}^k} f(T(x))dx &= \sum_{i=1}^{n} a_i \cdot \lambda(T^{-1}(A_i)) \\ &= \left|\det\left(T^{-1}\right)\right| \left(\sum_{i=1}^{n} a_i \cdot \lambda(A_i)\right) \\ &= \left|\det\left(T^{-1}\right)\right| \cdot \int_{\mathbb{R}^k} f(x)dx.\end{aligned}$$

Now let $f \in \mathcal{L}_+$ in general. Then, there exists a sequence $\{f_n\}_{n\in N_+}$ of simple non-negative $\mathcal{L}$-measurable functions increasing to f. $\{f_n \circ T\}_{n\in N_+}$ is then also a sequence of simple non-negative $\mathcal{L}$-measurable functions increasing to $f \circ T$; by repeated applications of the MCT, we have

$$\begin{aligned}\int_{\mathbb{R}^k} f(T(x))dx &= \lim_{n\to\infty}\int_{\mathbb{R}^k} f_n(T(x))dx \\ &= \left|\det\left(T^{-1}\right)\right|\cdot\left(\lim_{n\to\infty}\int_{\mathbb{R}^k} f_n(x)dx\right) \\ &= \left|\det\left(T^{-1}\right)\right|\cdot\int_{\mathbb{R}^k} f(x)dx.\end{aligned}$$

Finally, note from the product rule of determinants that, since $TT^{-1} = I_k$,

$$1 = \det\left(TT^{-1}\right) = \det(T)\det\left(T^{-1}\right),$$

so that $\det(T) = \frac{1}{\det(T^{-1})}$. Thus, for any $f \in \mathcal{L}_+$, we have

$$\int_{\mathbb{R}^k} f(x)dx = \frac{1}{|\det(T^{-1})|}\cdot\int_{\mathbb{R}^k} f(T(x))dx = |\det(T)|\cdot\int_{\mathbb{R}^k} f(T(x))dx.$$

Q.E.D.

Heuristically, given any $T \in \mathbb{R}^{k\times k}$, the determinant is the volume of the transformation $T[0,1]^k$ of the unit cube into a parallelogram. This is especially clear in the case of 2-dimensional space; letting $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $T[0,1]^2$ is the parallelogram with vertices at $(0,0)$, (a,c), (b,d) and $(a+b, c+d)$. The area of this parallelogram is then given as

$$\lambda(T[0,1]^2) = |\det(T)| = |ad - bc|.$$

Note that the angles of the parallelogram are given as the angle θ between the vectors (a,c) and (b,d), specifically

$$\theta = \arccos\left(\frac{|ab+cd|}{\sqrt{a^2+c^2}\sqrt{b^2+d^2}}\right)$$

and $\pi - \theta$. In the special case that (a,c) and (b,d) are orthogonal and are both of length 1, that is, $ab + cd = 0$ and $\sqrt{a^2+c^2} = \sqrt{b^2+d^2} = 1$, we can see that $\theta = \frac{\pi}{2}$ and the lengths of each edge of the parallelogram $T[0,1]^2$ equal 1. In other words, when (a,c) and (b,d) are orthogonal and of length 1, $T[0,1]^2$ is equivalent to a rotation of the unit square. Since (a,c) and (b,d) are orthogonal and of length 1 if and only if $T'T = I_2$, we can see that $T[0,1]^2$ rotates the unit square when T is an orthogonal matrix. This is why multiplication by an orthogonal matrix is often referred to as a rotation.

In a general euclidean k-space, the same idea holds. Given $T = (v_1, \cdots, v_k) \in \mathbb{R}^{k\times k}$, the transformation $T[0,1]^k$ is the parallelepiped with vertices at $\mathbf{0}$ and $v_{i_1} + \cdots + v_{i_m}$, where $1 \le m \le k$

and $1 \le i_1, \cdots, i_m \le k$ are distinct indices. Therefore, $T[0,1]^k$ has

$$\sum_{i=0}^{k} \binom{k}{i} = 2^k$$

vertices, the same number of vertices as $[0,1]^k$. If T is an orthogonal matrix, that is, if

$$T'T = \begin{pmatrix} v_1'v_1 & \cdots & v_1'v_k \\ \vdots & \ddots & \vdots \\ v_k'v_1 & \cdots & v_k'v_k \end{pmatrix} = I_k,$$

then the angle between any non-parallel edges, which are computed using the inner product of two vertices with distinct summands, is equal to 0, and the length of each edge is equal to 1. Therefore, $T[0,1]^k$ simply becomes a rotation of the unit cube $[0,1]^k$, so that a transformation of a set using an orthognoal matrix once again represents a rotation of the original set.

4.4 Continuity Properties of Measurable Functions

So far in this chapter, we have relied on linear functionals defined over the set of continuous compactly supported functions to construct and derive properties of Borel measures. This naturally leads one to question whether there exists a more direct relationship between continuous compactly supported functions and measurable functions. The answer to this is precisely the content of Lusin's theorem, stated below:

Theorem 4.14 (Lusin's Theorem)
Let (E,τ) be a locally compact Hausdorff space and $\mathcal{B}(E,\tau)$ the corresponding Borel σ-algebra. Let $(E,\mathcal{E},\mu)$ be a measure space with properties $i)$ to $iv)$ of the Riesz representation theorem:

i) $\mu(K) < +\infty$ for any compact $K \subset E$

ii) Any $\mathcal{E}$-measurable set is outer regular; for any $A \in \mathcal{E}$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

iii) Any $\mathcal{E}$-measurable set with finite measure or any open set is inner regular; for any $A \in \mathcal{E}$ such that $\mu(A) < +\infty$ or $A \in \tau$,

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

iv) $(E,\mathcal{E},\mu)$ is complete, and $\mathcal{E}$ contains all Borel sets.

Suppose f be a complex $\mathcal{E}$-measurable function such that $\mu(\{f \neq 0\}) < +\infty$. For any $\varepsilon > 0$, there exists a continuous compactly supported function $g \in C_c(E,\tau)$ such that

$$\mu(\{f \neq g\}) < \varepsilon.$$

Proof) Let $f \in \mathcal{E}_+$ be a non-negative measurable function taking values in $[0,1)$ with compact support K, and choose some $\varepsilon > 0$.
Let $\{s_n\}_{n\in N_+}$ a sequence of $\mathcal{E}$-measurable simple functions on E increasing to f, defined as in theorem 2.10. Define $\{f_n\}_{n\in N_+}$ as $f_1 = s_1$ and $f_n = s_n - s_{n-1} \geq 0$ for $n \geq 2$, $\{f_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable simple functions such that

$$f = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{i=1}^{n} f_i = \sum_{n=1}^{\infty} f_n.$$

Furthermore, for any $n \in N_+$, because $f(x) < 1$ for any $x \in E$,

$$\begin{aligned} s_n &= \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot I_{f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right)} + n \cdot I_{f^{-1}([n,+\infty])} \\ &= \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot I_{A_{nk}}, \end{aligned}$$

where $A_{nk} = f^{-1}\left(\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)\right)$ for $1 \le k \le n2^n$. Since

$$A_{nk} = f^{-1}\left(\left[\frac{2k-2}{2^{n+1}}, \frac{2k-1}{2^{n+1}}\right)\right) \cup f^{-1}\left(\left[\frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}}\right)\right) = A_{n+1,2k-1} \cup A_{n+1,2k},$$

we have

$$\begin{aligned} f_{n+1} = s_{n+1} - s_n &= \sum_{k=1}^{(n+1)2^{n+1}} \frac{k-1}{2^{n+1}} \cdot I_{A_{n+1,k}} - \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \cdot I_{A_{nk}} \\ &= \sum_{k=1}^{2^{n+1}} \frac{k-1}{2^{n+1}} \cdot I_{A_{n+1,k}} - \sum_{k=1}^{2^n} \frac{k-1}{2^n} \cdot I_{A_{nk}} \\ &= \sum_{k=1}^{2^n} \frac{2k-2}{2^{n+1}} \cdot I_{A_{n+1,2k-1}} + \sum_{k=1}^{2^n} \frac{2k-1}{2^{n+1}} \cdot I_{A_{n+1,2k}} \\ &\quad - \sum_{k=1}^{2^n} \frac{2k-2}{2^{n+1}} \cdot \left(I_{A_{n+1,2k-1}} + I_{A_{n+1,2k}}\right) \\ &= \sum_{k=1}^{2^n} \frac{1}{2^{n+1}} \cdot I_{A_{n+1,2k}} \\ &= \frac{1}{2^{n+1}} \left(\sum_{k=1}^{2^n} I_{A_{n+1,2k}}\right) = \frac{1}{2^{n+1}} I_{A_{n+1}}, \end{aligned}$$

where

$$A_{n+1} = f^{-1}\left(\bigcup_{k=1}^{2^n} \left[\frac{2k-1}{2^{n+1}}, \frac{2k}{2^{n+1}}\right)\right) \subset K$$

and we used the fact that

$$A_{n+1,k} = f^{-1}\left(\left[\frac{k-1}{2^{n+1}}, \frac{k}{2^{n+1}}\right)\right) = \emptyset$$

for any $1 \le k \le (n+1)2^{n+1}$ such that $k \ge 2^{n+1} + 1$ because $0 \le f < 1$ on E to justify the second equality.

As for f_1, we can see that

$$f_1 = s_1 = \sum_{k=1}^{2} \frac{k-1}{2} \cdot I_{A_{1k}} = \frac{1}{2} \cdot I_{A_1},$$

where we define $A_1 = A_{12} = f^{-1}\left(\left[\frac{1}{2}, 1\right)\right) \subset K$. It now follows that

$$f = \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot I_{A_n}$$

for a sequence $\{A_n\}_{n \in N_+}$ of $\mathcal{E}$-measurable sets such that $A_n \subset K$ for any $n \in N_+$.

Since K is a compact subset of the open set E, by theorem 1.14 it follows that there exists an open set V with compact closure $\overline{V}$ such taht $K \subset V \subset \overline{V} \subset E$.
For any $n \in N_+$, $\mu(A_n) \leq \mu(K) < +\infty$ by the monotonicity of μ. This means that A_n is both inner and outer regular, so that there exist an open set $G_n \in \tau$ and a compact set K_n such that $K_n \subset A_n \subset G_n$ and

$$\mu(G_n) < \mu(A_n) + \frac{\varepsilon}{2^{n+1}} \quad \text{and} \quad \mu(A_n) - \frac{\varepsilon}{2^{n+1}} < \mu(K_n).$$

Defining $V_n = G_n \cap V$, $V_n \in \tau$ and $A_n \subset V_n$ because $A_n \subset K \subset V$. It also holds that $\mu(V_n) \leq \mu(G_n) < \mu(A_n) + \frac{\varepsilon}{2^{n+1}}$. By implication,

$$\mu(V_n \setminus A_n) \leq \mu(V_n) - \mu(A_n) < \frac{\varepsilon}{2^{n+1}}$$
$$\mu(A_n \setminus K_n) \leq \mu(A_n) - \mu(K_n) < \frac{\varepsilon}{2^{n+1}},$$

and as such

$$\mu(V_n \setminus K_n) = \mu(V_n \setminus A_n) + \mu(A_n \setminus K_n) < \frac{\varepsilon}{2^n}.$$

By Urysohn's lemma, there exists an $h_n \in C_c(E, \tau)$ such that $K_n \prec h_n \prec V_n$, so that $h_n = 1$ on K_n, $\overline{\{h_n \neq 0\}} \subset V_n$ and $h_n \in [0,1]$ on E. Define the function

$$g = \sum_{n=1}^{\infty} \frac{1}{2^n} h_n;$$

because $\{\frac{1}{2^n} h_n\}_{n \in N_+}$ is a sequence of non-negative $\mathcal{E}$-measurable functions on E, by the MCT for series g is also a non-negative $\mathcal{E}$-measurable function. Furthermore, because

$$\left| \frac{1}{2^n} h_n \right| \leq \frac{1}{2^n}$$

on E for any $n \in N_+$, where $\sum_n \frac{1}{2^n} = 1 < +\infty$, by the Weierstrass M-test the sequence $\{g_n\}_{n \in N_+}$ of continuous functions defined as

$$g_n = \sum_{i=1}^{n} \frac{1}{2^i} h_i$$

for any $n \in N_+$ converges uniformly to g. Since continuity is preserved across uniform

limits, g is also a continuous function on E.
Additionally, because

$$K_n \subset A_n \subset V_n \subset V \subset \overline{V},$$

and

$$\{g \neq 0\} \subset \bigcup_n \{h_n \neq 0\} \subset \bigcup_n \overline{\{h_n \neq 0\}} \subset \bigcup_n V_n \subset \overline{V},$$

the support $\overline{\{g \neq 0\}}$ of g is a closed subset of the compact set $\overline{V}$ and thus compact. We have so far shown that $g \in C_c(E, \tau)$.

Now note that, because

$$\{f \neq g\} \subset \bigcup_n \{I_{A_n} \neq h_n\},$$

we have

$$\mu(\{f \neq g\}) \leq \sum_{n=1}^{\infty} \mu(\{I_{A_n} \neq h_n\}).$$

For any $n \in N_+$, $I_{A_n}(x) \neq h_n(x)$ implies $x \in V_n \setminus K_n$ (they are both 0 outside V_n and both 1 on K_n), so by monotonicity,

$$\mu(\{f \neq g\}) \leq \sum_{n=1}^{\infty} \mu(V_n \setminus K_n) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

We have thus shown that there existsa continuous compactly supported function g on E such that f and g differ on a set of measure no more than ε.

Suppose that f is now an $\mathcal{E}$-measurable non-negative function such that $|f| < M$ for some $M \in (0, +\infty)$ and $\mu(\{f \neq 0\}) < +\infty$. Defining $A = \{f \neq 0\}$, since $\mu(A) < +\infty$, A is inner regular, so that there exists a compact K such that $K \subset A$ and

$$\mu(A) - \frac{\varepsilon}{2} < \mu(K),$$

and by implication

$$\mu(A \setminus K) < \frac{\varepsilon}{2}.$$

Define

$$\bar{f} = \frac{1}{M} f \cdot I_K.$$

Then, $\bar{f}$ is a $\mathcal{E}$-measurable non-negative function taking values in $[0,1)$ with compact support K. By the preceding result, there exists a $\bar{g} \in C_c(E,\tau)$ such that

$$\mu(\{\bar{f} \neq \bar{g}\}) < \frac{\varepsilon}{4}.$$

Defining $g = M \cdot \bar{g}$, we can see that

$$\begin{aligned}\mu(\{f \neq g\} \cap A) &\leq \mu(\{f \neq g\} \cap A \cap K) + \mu(\{f \neq g\} \cap A \cap K^c) \\ &\leq \mu(\{\bar{f} \neq \bar{g}\}) + \mu(A \cap K^c) < \frac{3}{4}\varepsilon.\end{aligned}$$

Furthermore,

$$\mu(\{f \neq g\} \cap A^c) = \mu(\{g \neq 0\} \cap \{f = 0\}) \leq \mu(\{\bar{g} \neq \bar{f}\}) < \frac{\varepsilon}{4},$$

so we have

$$\mu(\{f \neq g\}) \leq \mu(\{f \neq g\} \cap A) + \mu(\{f \neq g\} \cap A^c) < \varepsilon.$$

Finally, let f be a non-negative real-valued $\mathcal{E}$-meausurable function with $\mu(\{f \neq 0\}) < +\infty$. Defining the sequence $\{B_n\}_{n \in N_+}$ of $\mathcal{E}$-measurable sets as

$$B_n = \{f \geq n\}$$

for any $n \in N_+$, since $B_{n+1} \subset B_n$ for any $n \in N_+$,

$$\bigcap_n B_n = \emptyset$$

because f is real-valued, and $\mu(B_1) = \mu(\{f > 1\}) \leq \mu(\{f \neq 0\}) < +\infty$, by the sequential continuity of measures

$$\lim_{n \to \infty} \mu(B_n) = \mu\left(\bigcap_n B_n\right) = 0.$$

This implies that there exists an $N \in N_+$ such that $\mu(B_N) < \frac{\varepsilon}{2}$.
Define $\bar{f} = f \cdot I_{B_N^c}$. Then, $\bar{f}$ is a $\mathcal{E}$ measurable non-negative function bounded above by N and with $\mu(\{\bar{f} \neq 0\}) \leq \mu(\{f \neq 0\}) < +\infty$. By the preceding result, there exists a $g \in C_c(E,\tau)$ such that

$$\mu(\{\bar{f} \neq g\}) < \frac{\varepsilon}{2}.$$

Therefore,

$$\begin{aligned}\mu(\{f \neq g\}) &\leq \mu(\{f \neq g\} \cap B_N^c) + \mu(\{f \neq g\} \cap B_N) \\ &\leq \mu(\{\bar{f} \neq g\}) + \mu(B_N) < \varepsilon.\end{aligned}$$

The claim now follows easily for the arbitrary complex valued $\mathcal{E}$-measurable function f such that $\mu(\{f \neq 0\}) < +\infty$. Since

$$f = (Re(f)^+ - Re(f)^-) + i \cdot (Im(f)^+ - Im(f)^-),$$

where $Re(f)^{\pm}, Im(f)^{\pm} \in \mathcal{E}_+$, and

$$\{f = 0\} \subset \{Re(f)^+ = Re(f)^-\} \cap \{Im(f)^+ = Im(f)^-\} = \{Re(f)^{\pm} = 0\} \cap \{Im(f)^{\pm} = 0\},$$

we have

$$\{Re(f)^{\pm} \neq 0\} \cup \{Im(f)^{\pm} \neq 0\} \subset \{f \neq 0\},$$

so that

$$\mu(\{Re(f)^{\pm} \neq 0\}), \mu(\{Im(f)^{\pm} \neq 0\}) \leq \mu(\{f \neq 0\}) < +\infty.$$

From the preceding result, it follows that there exist functions $g_1, g_2 \in C_c(E, \tau)$ and $h_1, h_2 \in C_c(E, \tau)$ such that

$$\mu(\{Re(f)^+ \neq g_1\}), \mu(\{Re(f)^- \neq g_2\}), \mu(\{Im(f)^+ \neq h_1\}), \mu(\{Im(f)^- \neq h_2\}) < \frac{\varepsilon}{4}.$$

Defiing $g = (g_1 - g_2) + i \cdot (h_1 - h_2) \in C_c(E, \tau)$, we can now see that

$$|f - g| \leq \left|Re(f)^+ - g_1\right| + |Re(f)^- - g_2| + \left|Im(f)^+ - h_1\right| + |Im(f)^- - h_2|,$$

which implies that

$$\{f \neq g\} \subset \{Re(f)^+ \neq g_1\} \cup \{Re(f)^- \neq g_2\} \cup \{Im(f)^+ \neq h_1\} \cup \{Im(f)^- \neq h_2\},$$

and as such

$$\mu(\{f \neq g\}) < \varepsilon.$$

Q.E.D.

Chapter 5

L^p Spaces

5.1 L^p Spaces as Vector Spaces

Let $(E, \mathcal{E}, \mu)$ be a measure space.
We start by noting that almost everywhere equivalence is an equivalence relation. To see this, first denote $f \sim g$ if $f = g$ a.e. $[\mu]$ for some $f, g : E \to \mathbb{C}$, where f, g are $\mathcal{E}$-measurable complex-valued:

i) (Reflexivity) For any $\mathcal{E}$-measurable numerical or complex function f, since $f(x) = f(x)$ for any $x \in E$, $f = f$ a.e. $[\mu]$ and $f \sim f$.

ii) (Symmetry) If $f \sim g$, then $g = f$ a.e. $[\mu]$ as well, so that $g \sim f$.

iii) (Transivitiy) If $f \sim g$ and $g \sim h$, then letting $A = \{f \neq h\}$, since $A \subset \{f \neq g\} \cup \{g \neq h\}$, by the subadditivity of measures

$$0 \leq \mu(A) \leq \mu(\{f \neq g\}) + \mu(\{g \neq h\}) = 0,$$

so that $\mu(A) = 0$. By definition, $f \sim h$.

In light of the above, for any $\mathcal{E}$-measurable numerical or complex function f we can define the equivalence class

$$[f]_\mu = \{g : E \to \mathbb{C} \mid g \text{ is } \mathcal{E}\text{-measurable and } f = g \text{ a.e. } [\mu]\},$$

or the set of all $\mathcal{E}$-measurable complex functions that are equal to f μ-almost everywhere. For any $\mathcal{E}$-measurable numerical or complex functions f, g, $f \sim g$ if and only if $[f]_\mu = [g]_\mu$; this follows easily from the fact that almost everywhere equivalence is an equivalence relation. Define $\mathcal{E}/\mathbb{C}$ as the collection of all equivalence classes of the above form.

Since the integrals of any μ-a.e. equivalent μ-integrable functions are equal, for any $p \in [1, +\infty)$

we can consider the following definition of the L^p space $L^p(\mathcal{E},\mu)$:

$$L^p(\mathcal{E},\mu) = \left\{[f]_\mu \in \mathcal{E}/\mathbb{C}\middle| \int_E |f|^p d\mu < +\infty\right\},$$

where $\int_E |f|^p d\mu = \int_E |g|^p d\mu$ for any $g \in [f]_\mu$ indicates that $\int_E |g|^p d\mu < +\infty$ for any $g \in [f]_\mu$ if $[f]_\mu \in L^p(\mathcal{E},\mu)$. In this sense, $L^p(\mathcal{E},\mu)$ is the collection of every equivalence class $[f]_\mu$ that is "integrable" in the pth power.
Note that, because p can be equal to 1, we have redefined $L^1(\mathcal{E},\mu)$ as a collection of equivalence classes instead of a set of functions.

$L^p(\mathcal{E},\mu)$ defined in this manner can be shown to be a vector space over the complex field. Define the addition and scalar multiplication of equivalence classes in $\mathcal{E}/\mathbb{C}$ by the addition and scalar multiplication of the representatives of those equivalence classes:

$$[f]_\mu + [g]_\mu = [f+g]_\mu$$
$$z \cdot [f]_\mu = [zf]_\mu$$

for any $[f]_\mu, [g]_\mu \in L^1(\mathcal{E},\mu)$ and $z \in \mathbb{C}$, where the notations on the right hand side are well defined because measurability of complex functions is preserved across addition and scalar multiplication. We now verify the axioms of a vector space.

1) **Closedness under Addition and Scalar Multiplication**
   Choose any $z \in \mathbb{C}$ and $[f]_\mu, [g]_\mu \in L^p(\mathcal{E},\mu)$. By definition,

$$z \cdot [f]_\mu + [g]_\mu = [zf+g]_\mu.$$

   To see that $|zf+g|^p$ is μ-integrable, note that, for any $x, y \in \mathbb{C}$,

$$|x+y|^p \leq (|x|+|y|)^p \leq 2^{p-1}(|x|^p + |y|^p),$$

   where the last inequality follows from the convexity of the function $\psi : (0,+\infty) \to (0,+\infty)$ defined as $\psi(a) = a^p$ for any $a \geq 0$.[1] By the monotonicity and linearity of integration,

$$\int_E |zf+g|^p d\mu \leq 2^{p-1} \cdot \int_E (|z|^p \cdot |f|^p + |g|^p)\, d\mu = 2^{p-1}|z|^p \cdot \int_E |f|^p d\mu + 2^{p-1} \cdot \int_E |g|^p d\mu < +\infty,$$

[1]To be more specific, the first derivative of ψ at any $a \in (0,+\infty)$ is $\psi'(a) = pa^{p-1}$, and its second derivative is $\psi''(a) = p(p-1)a^{p-2}$. Since $p \geq 1$, this meas that $\psi''(a) \geq 0$ for any $a > 0$, or that ψ is a convex function on $(0,+\infty)$. Therefore, for any $x, y \in \mathbb{C}$, if x, y are both not 0,

$$\left(\frac{|x|}{2} + \frac{|y|}{2}\right)^p = \psi\left(\frac{1}{2}|x| + \frac{1}{2}|y|\right) \leq \frac{1}{2}\psi(|x|) + \frac{1}{2}\psi(|y|) = \frac{1}{2}\left(|x|^p + |y|^p\right)$$

by convexity. Multiplying both sides by 2^p we get $(|x|+|y|)^p \leq 2^{p-1}(|x|^p + |y|^p)$.
On the other hand, if $x = 0$, then

$$(|x|+|y|)^p = |y|^p \leq 2^{p-1}|y|_p = 2^{p-1}(|x|^p + |y|^p).$$

The same holds when $y = 0$ as well.

where the last inequality follows because $[f]_\mu, [g]_\mu \in L^p(\mathcal{E},\mu)$. Therefore,

$$z \cdot [f]_\mu + [z]_\mu \in L^p(\mathcal{E},\mu),$$

and $L^p(\mathcal{E},\mu)$ is closed under addition and scalar multiplication.

2) **Commutativity of Addition**

For any $[f]_\mu, [g]_\mu \in L^p(\mathcal{E},\mu)$,

$$[f]_\mu + [g]_\mu = [f+g]_\mu = [g+f]_\mu = [g]_\mu + [f]_\mu,$$

where the middle equality follows from the commutatitivty of the pointwise addition of complex functions.

3) **Associativity of Addition**

For any $[f]_\mu, [g]_\mu, [h]_\mu \in L^p(\mathcal{E},\mu)$,

$$\begin{aligned}([f]_\mu + [g]_\mu) + [h]_\mu &= [f+g]_\mu + [h]_\mu = [(f+g)+h]_\mu \\ &= [f+(g+h)]_\mu = [f]_\mu + [g+h]_\mu = [f]_\mu + ([g]_\mu) + [h]_\mu,\end{aligned}$$

which follows by the associativity of the pointwise addition of complex functions.

4) **Existence of the Additive Identity**

Suppose $[f]_\mu \in L^p(\mathcal{E},\mu)$, and define $0_\mathcal{F}$ as the function on E defined as $0_\mathcal{F}(x) = 0$ for any $x \in E$. Clearly, the zero function is measurable and complex-valued, so $[0_\mathcal{F}]_\mu \in \mathcal{E}/\mathbb{C}$. Furthermore,

$$\int_E |0_\mathcal{F}|^p d\mu = 0 < +\infty,$$

so by definition $[0_\mathcal{F}]_\mu \in L^p(\mathcal{E},\mu)$. Now note that

$$[f]_\mu + [0_\mathcal{F}]_\mu = [f + 0_\mathcal{F}]_\mu = [f]_\mu$$

since $0_\mathcal{F}$ is the additive identity for complex functions defined on E.
By definition, $[0_\mathcal{F}]_\mu$ is the additive identity on $L^p(\mathcal{E},\mu)$.

5) **Existence of Additive Inverses**

Suppose $[f]_\mu \in L^p(\mathcal{E},\mu)$. Then, because $-f$ is a $\mathcal{E}$-measurable complex function such that

$$\int_E |-f|^p d\mu = \int_E |f|^p d\mu < +\infty,$$

by definition $[-f]_\mu \in L^p(\mathcal{E},\mu)$. Now note that

$$[f]_\mu + [-f]_\mu = [f - f]_\mu = [0_{\mathcal{F}}]_\mu,$$

since $-f$ is the additive inverse of f. It follows that $[-f]_\mu$ is the additive inverse of $[f]_\mu$.

6) **Commutativity of Scalar Multiplication**

   Suppose $[f]_\mu \in L^p(\mathcal{E},\mu)$ and $z,u \in \mathbb{C}$. Then,

$$z \cdot (u \cdot [f]_\mu) = z \cdot [uf]_\mu = [z(uf)]_\mu = [(zu)f]_\mu = (zu) \cdot [f]_\mu$$

   by the commutativity of scalar multiplication of complex functions defined on E.

7) **Scalar Multiplication by the Identity**

   For any $[f]_\mu \in L^p(\mathcal{E},\mu)$, becauase 1 is the multiplicative identity of the complex field,

$$1 \cdot [f]_\mu = [1 \cdot f]_\mu = [f]_\mu$$

   by the fact that $1 \cdot f = f$ on E.

8) **Distributive Law I**

   For any $[f]_\mu, [g]_\mu \in L^p(\mathcal{E},\mu)$ and $z \in \mathbb{C}$,

$$z \cdot ([f]_\mu + [g]_\mu) = z \cdot [f+g]_\mu = [z \cdot (f+g)]_\mu = [zf + zg]_\mu = z \cdot [f]_\mu + z \cdot [g]_\mu$$

   by the corresponding distributive law for complex functions defined on E.

9) **Distributive Law II**

   For any $[f]_\mu \in L^p(\mathcal{E},\mu)$ and $z,u \in \mathbb{C}$,

$$(z+u) \cdot [f]_\mu = [(z+u) \cdot f]_\mu = [zf + uf]_\mu = z \cdot [f]_\mu + u \cdot [f]_\mu$$

   by the corresponding distributive law for complex functions defined on E.

Since $L^p(\mathcal{E},\mu)$ satisfies all the vector space axioms, it is indeed a vector space over the complex field. The whole host of vector space results thus also applies to the vector space $(L^p(\mathcal{E},\mu),\mathbb{C})$. Note, however, that in general $L^p(\mathcal{E},\mu)$ is an infinite-dimensional space.

5.2 L^p Spaces as Normed Vector Spaces

To facilitate the exposition of this section, we first prove some results related to convex functions defined on the real line and present a few useful inequalities.

5.2.1 Convex Functions and Jensen's Inequality

For any open interval (a,b) where $-\infty \leq a < b \leq +\infty$, the real valued function $\varphi : (a,b) \to \mathbb{R}$ is said to be convex if, for any $x,y \in (a,b)$ and $t \in [0,1]$,

$$\varphi(tx+(1-t)y) \leq t\varphi(x)+(1-t)\varphi(y).$$

For any $a < x < y < z < b$, letting $t = \frac{y-x}{z-x} \in (0,1)$, we thus have

$$\varphi(tx+(1-t)z) \leq t\varphi(x)+(1-t)\varphi(z);$$

because $tx+(1-t)z = z-t(z-x) = y$, the above inequality becomes

$$\frac{\varphi(y)-\varphi(x)}{y-x} \leq \frac{\varphi(z)-\varphi(x)}{z-x}.$$

Similarly, letting $t = \frac{z-y}{z-x} \in (0,1)$,

$$\varphi(tx+(1-t)z) \leq t\varphi(x)+(1-t)\varphi(z),$$

and since $tx+(1-t)z = z+t(x-z) = y$, the above inequality becomes

$$\varphi(y) \leq \varphi(z) - \frac{z-y}{z-x}\left(\varphi(z)-\varphi(x)\right),$$

or

$$\frac{\varphi(z)-\varphi(x)}{z-x} \leq \frac{\varphi(z)-\varphi(y)}{z-y}.$$

Putting the two inequalities together yields

$$\frac{\varphi(y)-\varphi(x)}{y-x} \leq \frac{\varphi(z)-\varphi(y)}{z-y}.$$

Using the properties shown above, we can prove that any convex function $\varphi : (a,b) \to \mathbb{R}$ is continuous on (a,b):

Lemma 5.1 (Continuity of Convex Functions)
Every convex function defined on a real open interval is continuous with respect to the euclidean metric on $\mathbb{R}$.

Proof) Choose any $x \in (a,b)$, and a sequence $\{x_n\}_{n \in N_+}$ in (a,b) that increases to x but does not contain x. Then, for any $n \in N_+$, since $x_1 \leq x_n < x$, we have

$$\frac{\varphi(x)-\varphi(x_1)}{x-x_1} \leq \frac{\varphi(x)-\varphi(x_n)}{x-x_n}.$$

Rearranging terms yields

$$\varphi(x_n) \leq \frac{x_n - x_1}{x-x_1}\varphi(x) + \frac{x-x_n}{x-x_1}\varphi(x_1),$$

and because the right hand side converges to $\varphi(x)$ as $n \to \infty$, we have

$$\limsup_{n\to\infty} \varphi(x_n) \leq \varphi(x).$$

Similarly, letting $x < y < b$,

$$\frac{\varphi(x)-\varphi(x_n)}{x-x_n} \leq \frac{\varphi(y)-\varphi(x_n)}{y-x_n},$$

which can be rearranged into

$$\varphi(x) \leq \frac{x-x_n}{y-x_n}\varphi(y) + \frac{y-x}{y-x_n}\varphi(x_n).$$

Taking the limit inferior on both sides yields

$$\varphi(x) \leq \liminf_{n\to\infty} \varphi(x_n).$$

Therefore,

$$\limsup_{n\to\infty} \varphi(x_n) = \liminf_{n\to\infty} \varphi(x_n) = \varphi(x) = \lim_{n\to\infty} \varphi(x_n),$$

and because this holds for any sequence increasing to x,

$$\varphi(x) = \lim_{y \to x^-} \varphi(y).$$

Now consider a sequence $\{x_n\}_{n \in N_+}$ in (a,b) decreasing to x but not containing x. Then, for any $n \in N_+$, $x < x_n \leq x_1$, and

$$\frac{\varphi(x_n)-\varphi(x)}{x_n-x} \leq \frac{\varphi(x_1)-\varphi(x)}{x_1-x}.$$

Rearranging terms yields

$$\varphi(x_n) \le \frac{x_1 - x_n}{x_1 - x}\varphi(x) + \frac{x_n - x}{x_1 - x}\varphi(x_1),$$

and taking the limit superior on both sides, we have

$$\limsup_{n\to\infty} \varphi(x_n) \le \varphi(x).$$

Letting $a < y < x$, since $y < x < x_n$ for any $n \in N_+$, we can see that

$$\frac{\varphi(x) - \varphi(y)}{x - y} \le \frac{\varphi(x_n) - \varphi(y)}{x_n - y}.$$

Rearranging terms as above, we end up with

$$\varphi(x) + \frac{x - x_n}{x_n - y}\varphi(y) \le \frac{x - y}{x_n - y}\varphi(x_n),$$

so taking limit inferiors on oboth sides,

$$\varphi(x) \le \liminf_{n\to\infty} \varphi(x_n).$$

As in the case for the left hand side limit, this implies that

$$\lim_{n\to\infty} \varphi(x_n) = \varphi(x)$$

and therefore that

$$\varphi(x) = \lim_{y\to x^+} \varphi(y).$$

Since $\varphi(x) = \lim_{y\to x^+} \varphi(y) = \lim_{y\to x^-} \varphi(y)$, it follows that

$$\varphi(x) = \lim_{y\to x} \varphi(y),$$

and therefore φ is continuous at x.[2]

Q.E.D.

[2] To make the proof truly rigorous, we must show that, if $\varphi(x_n) \to \varphi(x)$ for any sequence $\{x_n\}_{n\in N_+}$ decreasing or increasing to x, then $\varphi(y) \to \varphi(x)$ as $y \to x$. This can be shown by proving the contrapositive.
Specifically, suppose that $\varphi(y)$ does not converge to $\varphi(x)$ as $y \to x$. Then, there exists a $\varepsilon > 0$ such that, for any $\delta > 0$ there exists a $y \in (a,b)$ such that $|y - x| < \delta$ and $|\varphi(y) - \varphi(x)| \ge \varepsilon$. This means that, for any $n \in N_+$, we can choose an $x_n \in (a,b)$ such that $|x_n - x| < \frac{1}{n}$ and $|\varphi(x_n) - \varphi(x)| \ge \varepsilon$. Clearly, $\{x_n\}_{n\in N_+}$ is a sequence in (a,b) converging to x, and $x_n \ne x$ for any $n \in N_+$.
Suppose that there are only finitely many $n \in N_+$ such that $x_n < x$; this means that there are infinitely many $n \in N_+$ such that $x < x_n$, since $x_n \ne x$ for all n. A similar argument shows that there must either be infinitely many $n \in N_+$ such that $x_n > x$ or infinitely many $n \in N_+$ such that $x_n < x$.
Without loss of generality, suppose that $x < x_n$ for infinitely many $n \in N_+$. Then, we can construct a subsequence $\{x_{n_k}\}_{k\in N_+}$ of $\{x_n\}_{n\in N_+}$ that decreases to x, using the fact that $x_n \to x$ as $n \to \infty$. By design, $|\varphi(x_{n_k}) - \varphi(x)| \ge \varepsilon$, so $\varphi(x_{n_k})$ does not converge to $\varphi(x)$ as $k \to \infty$.
We have just shown that there must exist a sequence $\{x_n\}_{n\in N_+}$ in (a,b) that does not contain x and either increases or decreases to x such that $\varphi(x_n)$ does not converge to $\varphi(x)$. The claim then follows by contraposition.

Another useful property of convex functions is that they are bounded below. If the convex function in question is twice differentiable, then this is very easy to show, since their second derivative must always be increasing and thus they have a global minimum. We can show that this is true even if the function is not differentiable:

Lemma 5.2 (Boundedness of Convex Functions)
Every convex function defined on a real open interval is bounded below.

Proof) Let $\varphi : (a,b) \to \mathbb{R}$ be a convex function, where $-\infty \le a < b \le +\infty$. Suppose that φ is not bounded below. Then, there exists a sequence $\{x_n\}_{n \in N_+}$ on (a,b) such that

$$\varphi(x_n) \searrow -\infty.$$

Since $\{x_n\}_{n \in N_+}$ takes values in the compact set $[a,b]$, by the equivalence of sequential compactness and compactness there exists a subsequence $\{x_{n_k}\}_{k \in N_+}$ of $\{x_n\}_{n \in N_+}$ that converges to some $x \in [a,b]$. If $x \in (a,b)$, then by the continuity of φ,

$$\varphi(x) = \lim_{n \to \infty} \varphi(x_n) = -\infty,$$

which contradicts the fact that $\varphi(x) \in \mathbb{R}$. Therefore $x = a$ or $x = b$.

Without loss of generality, suppose $x = a$. Because $\{x_{n_k}\}_{k \in N_+}$ is a sequence converging to a, for $\varepsilon > 0$ chosen so that $a + 2\varepsilon < b$, there exists an $N \in N_+$ such that

$$|x_{n_k} - a| = x_{n_k} - a < \varepsilon$$

for any $k \ge N$. Denoting $a < y = a + \varepsilon < a + 2\varepsilon = z < b$, this implies that, for any $k \ge N$,

$$a < x_{n_k} < y < z < b.$$

Defining $t_k = \frac{z-y}{z-x_{n_k}} \in (0,1)$ for any $k \in N_+$, $t_k \to \frac{z-y}{z-a} = t \in (0,1)$ as $k \to \infty$, and by the convexity of φ,

$$\varphi(y) = \varphi(t_k x_{n_k} + (1-t_k)z) \le t_k \varphi(x_{n_k}) + (1-t_k)\varphi(z).$$

Sending $k \to \infty$ on both sides yields

$$\varphi(y) \le t \cdot (-\infty) + (1-t)\varphi(z) = -\infty,$$

which implis that $\varphi(y) = -\infty$. This is a contradiction because φ is real-valued, and as such, φ must be bounded below.
Q.E.D.

Let $(E, \mathcal{E}, \mu)$ be a probability space. We will now prove a famous inequality involving convex functions:

Theorem 5.3 (Jensen's Inequality)
Let $(E, \mathcal{E}, \mu)$ be a probability space, $\varphi : (a,b) \to \mathbb{R}$ a convex function for some $-\infty \leq a < b \leq \infty$, and f a real-valued $\mathcal{E}$-measurable function that takes values in (a,b). Then, f is μ-integrable, $\int_E f d\mu \in (a,b)$ and $\varphi \circ f$ is also an $\mathcal{E}$-measurable real valued function.
Morever, the integral of $\varphi \circ f$ with respect to μ exists in the extended sense, and

$$\varphi\left(\int_E f d\mu\right) \leq \int_E (\varphi \circ f)\, d\mu.$$

Proof) There are many items to prove in this theorem: hence we proceed in steps.

Step 1: The Measurability of $\varphi \circ f$

Because φ is convex, it is continuous on (a,b). Before using the results that continuous functions are measurable and the composition of measurable functions is measurable, note that φ is defined on a subset (a,b) of $\mathbb{R}$, and that we have not defined a topology on (a,b). This means that we cannot conclude that φ is measurable based on its continuity, and as such that we cannot rely on previous results at the current stage.

Instead, recall the results from PMA (Baby Rudin). In chapter 4 of that textbook, we learned that, for any set B open relative to $\mathbb{R}$ (we now know that this basically means that B is an element of the metric topology $\tau_{\mathbb{R}}$ on $\mathbb{R}$), $\varphi^{-1}(B)$ must be a set that is open relative to (a,b) by the continuity of φ. Chapter 2 of the textbook also tells us that a set open relative to (a,b) is the intersection of (a,b) and a set open relative to $\mathbb{R}$, so there exists an $A \in \tau_{\mathbb{R}}$ such that $\varphi^{-1}(B) = A \cap (a,b)$. Since $A \cap (a,b)$ is itself an open set in $\mathbb{R}$, we can see that $\varphi^{-1}(B) \in \tau_{\mathbb{R}}$.
This then implies that

$$(\varphi \circ f)^{-1}(B) = f^{-1}(\varphi^{-1}(B)) \in \mathcal{E}$$

because f is measurable and $\tau_{\mathbb{R}} \subset \mathcal{B}(\mathbb{R})$. Finally, $\tau_{\mathbb{R}}$ is a set that generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, so the fact that $(\varphi \circ f)^{-1}(B) \in \mathcal{E}$ for any $B \in \tau_{\mathbb{R}}$ indicates that $\varphi \circ f$ is a $\mathcal{E}$-measurable real function.

Step 2: $\int_E f d\mu$ is contained in (a,b)

For any $x \in E$, $a < f(x) < b$; because μ is a probability measure, constant functions are always μ-integrable, and therefore, by the monotonicity of integration,

$$a = a \cdot \mu(E) = \int_E a d\mu \leq \int_E f d\mu \leq \int_E b d\mu = b \cdot \mu(E) = b,$$

which tells us that $\int_E f d\mu \in [a,b]$ and thus that f is μ-integrable.
Suppose that $\int_E f d\mu = a$. Then, by the linearity of integration

$$\int_E g d\mu = \int_E f d\mu - a = 0.$$

Since $g = f - a > 0$ on E because $f(x) > a$ for any $x \in E$, the vanishing property for non-negative functions tells us that $g = 0$ a.e. $[\mu]$. This, however, contradicts the fact that $g > 0$ everywhere on E, so $a < \int_E f d\mu$. It can be shown through a similar process that $\int_E f d\mu < b$, so $\int_E f d\mu \in (a,b)$.

Step 3: An Auxiliary Inequality for Convex Functions

To show the main inequality, we exploit a special property of convex function. Recall that φ, being a convex function, satisfies

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(y)}{z - y}$$

for any $a < x < y < z < b$. Fixing y, define

$$\beta_y = \sup_{x \in (a,y)} \frac{\varphi(y) - \varphi(x)}{y - x},$$

which exists in $\mathbb{R}$ because the set $\{\frac{\varphi(y)-\varphi(x)}{y-x} \mid x \in (a,y)\}$ is bounded above by $\frac{\varphi(z)-\varphi(y)}{z-y} \in \mathbb{R}$ for any $y < z < b$ and the real line possess the least upper bound property. By the definition of the supremum as the least upper bound, we have

$$\beta_y \leq \frac{\varphi(z) - \varphi(y)}{z - y}.$$

for any $z \in (y,b)$. Multiplying both sides by $z - y$ now yields the inequality

$$\varphi(z) \geq \varphi(y) + \beta_y(z - y)$$

for any $z \in (y,b)$. For any $a < x < y$,

$$\frac{\varphi(y) - \varphi(x)}{y - x} \leq \beta_y$$

by definition of the supremum, and as such

$$\varphi(x) \geq \varphi(y) + \beta_y(x-y).$$

Therefore, for any $z \in (a,b)$, we have

$$\varphi(z) \geq \varphi(y) + \beta_y(z-y),$$

where β_y is the supremum defined above.

Now put $y = \int_E f d\mu \in (a,b)$; then, by the preceding result, for any $x \in E$

$$(\varphi \circ f)(x) = \varphi(f(x)) \geq \varphi(y) + \beta_y(f(x) - y),$$

since $f(x) \in (a,b)$. Therefore,

$$\beta_y \cdot f + k \leq \varphi \circ f$$

on E, where we put $k = \varphi(y) - \beta_y \cdot y \in \mathbb{R}$. Note that $\beta_y \cdot f + k \in L^1(\mathcal{E},\mu)$ because f and k are μ-integrable (μ is a probability measure).

Step 5: Jensen's Inequality when $\varphi \circ f$ is Integrable

Suppose that $\varphi \circ f$ is μ-integrable. Then, by the monotonicity and linearity of integration,

$$\beta_y \cdot \int_E f d\mu + k = \int_E (\beta_y \cdot f + k) d\mu \leq \int_E (\varphi \circ f)\, d\mu,$$

and because $k = \varphi(y) - \beta_y \cdot y = \varphi\left(\int_E f d\mu\right) - \beta_y \cdot \int_E f d\mu$, we have

$$\varphi\left(\int_E f d\mu\right) \leq \int_E (\varphi \circ f)\, d\mu.$$

Step 6: Jensen's Inequality when $\varphi \circ f$ is not Integrable

On the other hand, suppose that $\varphi \circ f$ is not μ-integrable, that is, that either $\int_E (\varphi \circ f)^+ \, d\mu = +\infty$ or $\int_E (\varphi \circ f)^- \, d\mu = +\infty$. Since φ is convex and a convex function is bounded below, letting $M \leq \varphi(x)$ for any $x \in (a,b)$ and some $-\infty < M < 0$, $M \leq \min(\varphi, 0)$ and thus

$\varphi^- \leq -M$, which implies that

$$int_E \, (\varphi \circ f)^- \, d\mu \leq -M < +\infty.$$

Therefore, if $\varphi \circ f$ is not μ-integrable, then it must be the case that

$$\int_E (\varphi \circ f)^+ \, d\mu = +\infty, \quad \text{while} \quad \int_E (\varphi \circ f)^- \, d\mu < +\infty.$$

This means that the integral of $\varphi \circ f$ is defined over E in the extended sense and equals

$$\int_E (\varphi \circ f) \, d\mu = \int_E (\varphi \circ f)^+ \, d\mu - \int_E (\varphi \circ f)^- \, d\mu = +\infty.$$

It is now trivial that

$$\varphi \left(\int_E f d\mu \right) \leq +\infty = \int_E (\varphi \circ f) \, d\mu.$$

Q.E.D.

Jensen's inequality, in the form presented above, says that the convex function of a measurable function is also measurable, and that the convex function of the integral is always less than or equal to the integral of the convex function, where the latter integral always exists in $(-\infty, +\infty]$. Thus, it furnishes us with the measurability and (in the extended sense) the existence of the integral of convex functions of measurable functions, as well as a lower bound to that integral.

If the integral above is interpreted as an expectation with respect to the probability measure μ, then it yields the version of Jensen's inequality often encountered in probability theory.

5.2.2 Hölder's Inequality and Minkowski's Inequality

Here we present two inequalities crucial to establishing the algebraic propreties of L^p spaces. Hölder's inequality is a generalized version of the Cauchy-Schwarz inequality, and Minkowski's inequality is the generalization of the triangle inequality.
We first present an inequality that will prove useful, not only in the proof of Hölder and Minkowski's Inequalities, but in general. It is proved in this text as a corollary of Jensen's inequality, but there are more classical ways to prove it (see: exercise 10 of chapter 6 in PMA):

Lemma 5.4 (Young's Inequality)
For any $x, y \in [0, +\infty]$ and $p, q \in (1, +\infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q.$$

When $x, y \in [0, +\infty)$, then the above inequality is an equality if and only if $x^p = y^q$.

Proof) The case for $x = 0$ and $x = +\infty$ are obvious, as is the corresponding case for y. If $x = 0$ and the inequality holds as an equality, then $0 = \frac{1}{q}y^q$ and thus $y = 0$, implying $x^p = y^q = 0$. The same goes for when $y = 0$.

Now let $x, y \in (0, +\infty)$. The claim follows easily in this case as well:

$$\begin{aligned} xy = e^{\log(xy)} &= e^{\log(x)+\log(y)} \\ &= e^{\frac{1}{p}(p\log(x))+\frac{1}{q}(q\log(y))} \\ &\leq \frac{1}{p}e^{p\log(x)} + \frac{1}{q}e^{q\log(y)} && \text{(Convexity of the exponential function)} \\ &= \frac{1}{p}x^p + \frac{1}{q}y^q. \end{aligned}$$

In addition, note that

$$\begin{aligned} xy &= e^{\frac{1}{p}(p\log(x))+\frac{1}{q}(q\log(y))} \\ &= \frac{1}{p}x^p + \frac{1}{q}y^q = \frac{1}{p}e^{p\log(x)} + \frac{1}{q}e^{q\log(y)}, \end{aligned}$$

if and only if $p\log(x) = q\log(y)$, or $x^p = y^q$. Sufficiency is obvious, and necessity follows since, if $p\log(x) \neq q\log(y)$, then

$$e^{\frac{1}{p}(p\log(x))+\frac{1}{q}(q\log(y))} < \frac{1}{p}e^{p\log(x)} + \frac{1}{q}e^{q\log(y)}$$

by $\frac{1}{p} \in (0, 1)$ and the strict convexity of the exponential function.
Q.E.D.

Theorem 5.5 (Hölder and Minkowski's Inequalities)
Let $(E,\mathcal{E},\mu)$ be a measure space, and f,g non-negative $\mathcal{E}$-measurable functions. Then, the following inequalities hold:

- **(Hölder's Inequality)** For any $p,q \in (1,+\infty)$ such that $\frac{1}{p}+\frac{1}{q}=1$,

$$\int_E (fg)d\mu \leq \left(\int_E f^p d\mu\right)^{\frac{1}{p}} \left(\int_E g^q d\mu\right)^{\frac{1}{q}}.$$

  If $[f]_\mu,[g]_\mu \in L^p(\mathcal{E},\mu)$, then the inequality is an equality if and only if there exist $a,b \in [0,+\infty)$ such that $a>0$ or $b>0$ and $a\cdot f^p = b\cdot g^q$ a.e. $[\mu]$.

- **(Minkowski's Inequality)** For any $p \in [1,+\infty)$,

$$\left(\int_E (f+g)^p d\mu\right)^{\frac{1}{p}} \leq \left(\int_E f^p d\mu\right)^{\frac{1}{p}} + \left(\int_E g^p d\mu\right)^{\frac{1}{p}}.$$

Proof) We also proceed in steps with the proof of this theorem.

Step 1: Proving Hölder's Inequality

Let $p,q \in (1,+\infty)$ satisfy $\frac{1}{p}+\frac{1}{q}=1$. Define $A=(\int_E f^p d\mu)^{\frac{1}{p}}$ and $B=(\int_E g^q d\mu)^{\frac{1}{q}}$, which are both well-defined in $[0,+\infty]$ because the integrals are of non-negative measurable functions.
If $A=+\infty$, then because the right hand side of the inequality becomes $+\infty$, it holds trivially; the same goes for the case where $B=+\infty$.
If $A=0$, then by the vanishing property of non-negative functions, $f^p=0$ and thus $f=0$ a.e. $[\mu]$, meaning that $fg=0$ a.e. $[\mu]$ and

$$\int_E (fg)d\mu = 0 = A\cdot B.$$

The same goes for the case where $B=0$.

It remains to prove the inequality for the case where $A,B \in (0,+\infty)$. To this end, define the functions $F,G : \mathbb{E} \to [0,+\infty]$ as

$$F(x)=\frac{f(x)}{A} \quad \text{and} \quad G(x)=\frac{g(x)}{B}$$

for any $x \in E$. Because f,g are measurable non-negative functions, so are F,G, and they satisfy

$$\int_E F^p d\mu = \frac{1}{A^p}\cdot\int_E f^p d\mu = 1 \quad \text{and} \quad \int_E G^q d\mu = \frac{1}{B^q}\cdot\int_E g^q d\mu = 1$$

by the linearity of integration.

For any $x \in E$, $F(x), G(x) \in [0, +\infty]$, so by Young's inequality,

$$F(x)G(x) \leq \frac{1}{p}F(x)^p + \frac{1}{q}G(x)^q.$$

The functions fg, FG and $\frac{1}{p}F^p + \frac{1}{q}G^q$ are both measurable non-negative functions, so their integrals exist, and by the monotonicity and linearity of integration,

$$\frac{1}{AB}\int_E (fg)d\mu = \int_E (FG)d\mu \leq \frac{1}{p}\int_E F^p d\mu + \frac{1}{q}\int_E G^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

Thus, multiplying both sides by $AB > 0$ yields

$$\int_E (fg)d\mu \leq AB = \left(\int_E f^p d\mu\right)^{\frac{1}{p}} \left(\int_E g^q d\mu\right)^{\frac{1}{q}},$$

which is the desired inequality.

Step 2: Necessary Conditions for Hölder's Inequality to hold as an Equality

To derive the necessary conditions when the inequality is an equality in the case where $[f]_\mu, [g]_\mu \in L^p(\mathcal{E}, \mu)$, first assume that $A = 0$ or $B = 0$. Without loss of generality, suppose that $A = 0$. Then, because $f = 0$ a.e. $[\mu]$, the inequality always holds as an equality and $1 \cdot f^p = 0 \cdot g^q$ a.e. $[\mu]$.

Now assume that $A, B \in (0, +\infty)$. Because f, g are now real non-negative valued, F, G defined above must take values in $[0, +\infty)$.
We showed above that $FG \leq \frac{1}{p}F^p + \frac{1}{q}G^q$, so defining

$$h = \frac{1}{p}F^p + \frac{1}{q}G^q - FG,$$

which is well-defined because F, G are real valued, h is a measurable non-negative function.
Suppose that Hölder's inequality holds as an equality, that is, $\int_E (fg)d\mu = AB$. Dividing both sides by AB reveals that this is equivalent to

$$\int_E FG d\mu = 1 = \frac{1}{p} + \frac{1}{q} = \int_E \left(\frac{1}{p}F^p + \frac{1}{q}G^q\right) d\mu,$$

and by the linearity of integration,

$$\int_E h d\mu = \int_E \left(\frac{1}{p}F^p + \frac{1}{q}G^q - FG\right) d\mu = 0.$$

By the vanishing property of non-negative functions, this means that $h = 0$ a.e. $[\mu]$.

For any $x \in E$ such that $h(x) = 0$,

$$F(x)G(x) = \frac{1}{p}F(x)^p + \frac{1}{q}G(x)^q.$$

Here, $F(x), G(x) \in [0, +\infty)$, so by the necessary condition for Young's inequality to be an equality, we have $F(x)^p = G(x)^q$.
This holds for any $x \in E$ such that $h(x) = 0$, so $F^p = G^q$, or $B \cdot f^p = A \cdot g^q$, a.e. $[\mu]$.

Step 3: Sufficient Conditions for Hölder's Inequality to hold as an Equality

Finally, it remains to show that $a \cdot f^p = b \cdot g^q$ a.e. $[\mu]$ for some $a, b \in [0, +\infty)$ such that $a > 0$ or $b > 0$ is a sufficient condition for the inequality to hold as an equality when $[f]_\mu, [g]_\mu \in L^p(\mathcal{E}, \mu)$. Suppose without loss of generality that $a > 0$. Then, $f = \left(\frac{b}{a}\right)^{\frac{1}{p}} g^{\frac{q}{p}}$, which implies that

$$\int_E f^p d\mu = \frac{b}{a} \cdot \int_E g^q d\mu,$$
$$\left(\int_E f^p d\mu\right)^{\frac{1}{p}} = \left(\frac{b}{a}\right)^{\frac{1}{p}} \left(\int_E g^q d\mu\right)^{\frac{1}{p}}$$

and

$$\begin{aligned}\int_E (fg) d\mu = \left(\frac{b}{a}\right)^{\frac{1}{p}} \cdot \int_E g^{\frac{p+q}{p}} d\mu &= \left(\frac{b}{a}\right)^{\frac{1}{p}} \cdot \int_E g^q d\mu \\ &= \left(\frac{b}{a}\right)^{\frac{1}{p}} \cdot \left(\int_E g^q d\mu\right)^{\frac{1}{p}} \left(\int_E g^q d\mu\right)^{\frac{1}{q}} \\ &= \left(\int_E f^p d\mu\right)^{\frac{1}{p}} \left(\int_E g^q d\mu\right)^{\frac{1}{q}}.\end{aligned}$$

Therefore, the stated condition is also sufficient for Hölder's inequality to hold as an equality.

Step 4: Proving Minkowski's Inequality

Minkowski's inequality is much easier to prove using Hölder's inequality. Let $p \in [1, +\infty)$. If $p = 1$, then the result follows from the linearity of integration:

$$\int_E (f+g) d\mu = \int_E f d\mu + \int_E g d\mu.$$

Now let $p \in (1, +\infty)$. If $\int_E f^p d\mu = +\infty$ or $\int_E g^p d\mu = +\infty$, then the inequality holds trivially, so let $\int_E f^p d\mu, \int_E g^p d\mu \in [0, +\infty)$. In this case, due to the closedness of L^p spaces under addition and scalar multiplication, $\int_E (f+g)^p d\mu < +\infty$ as well.
If $\int_E (f+g)^p d\mu = 0$, then the inequality also holds trivially, so assume $\int_E (f+g)^p d\mu \in (0, +\infty)$. Note that

$$(f+g)^p = (f+g)(f+g)^{p-1} = f \cdot (f+g)^{p-1} + g \cdot (f+g)^{p-1}.$$

Letting $r = \frac{p}{p-1}$, because $\frac{1}{r} + \frac{1}{p} = \frac{p-1}{p} + \frac{1}{p} = 1$, by Hölder's inequality we have

$$\begin{aligned} \int_E \left(f \cdot (f+g)^{p-1} \right) d\mu &\leq \left(\int_E f^p d\mu \right)^{\frac{1}{p}} \left(\int_E (f+g)^{(p-1)r} d\mu \right)^{\frac{1}{r}} \\ &= \left(\int_E f^p d\mu \right)^{\frac{1}{p}} \left(\int_E (f+g)^p d\mu \right)^{1-\frac{1}{p}}, \end{aligned}$$

and likewise,

$$\int_E \left(g \cdot (f+g)^{p-1} \right) d\mu \leq \left(\int_E g^p d\mu \right)^{\frac{1}{p}} \left(\int_E (f+g)^p d\mu \right)^{1-\frac{1}{p}}.$$

By the linearity of integration, it follows that

$$\begin{aligned} \int_E (f+g)^p d\mu &= \int_E \left(f \cdot (f+g)^{p-1} \right) d\mu + \int_E \left(g \cdot (f+g)^{p-1} \right) d\mu \\ &\leq \left(\int_E f^p d\mu \right)^{\frac{1}{p}} \left(\int_E (f+g)^p d\mu \right)^{1-\frac{1}{p}} + \left(\int_E g^p d\mu \right)^{\frac{1}{p}} \left(\int_E (f+g)^p d\mu \right)^{1-\frac{1}{p}} \\ &= \left[\left(\int_E f^p d\mu \right)^{\frac{1}{p}} + \left(\int_E g^p d\mu \right)^{\frac{1}{p}} \right] \left(\int_E (f+g)^p d\mu \right)^{1-\frac{1}{p}}. \end{aligned}$$

Since $\int_E (f+g)^p d\mu \in (0, +\infty)$, we can divide both sides by $\left(\int_E (f+g)^p d\mu\right)^{1-\frac{1}{p}}$ to obatin the desired result

$$\left(\int_E (f+g)^p d\mu \right)^{\frac{1}{p}} \leq \left(\int_E f^p d\mu \right)^{\frac{1}{p}} + \left(\int_E g^p d\mu \right)^{\frac{1}{p}}.$$

Q.E.D.

5.2.3 The L^p Norm

Let $(E, \mathcal{E}, \mu)$ be a measure space. We proved in section 1 that, for any $p \in [1, +\infty)$, the space $L^p(\mathcal{E}, \mu)$ of equivalence classes is a vector space over the complex field. We now define a norm on this vector space, so that it may be rendered a normed vector space.

Define the function $\|\cdot\|_p$ as

$$\|[f]_\mu\|_p = \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}}$$

for any $[f]_\mu \in \mathcal{E}/\mathbb{C}$. Clearly, $\|[f]_\mu\|_p \in \mathbb{R}_+$ if $[f]_\mu \in L^p(\mathcal{E}, \mu)$. In addition, we can show the following properties of $\|\cdot\|_p$:

- Suppose $\|[f]_\mu\|_p = 0$ for some $[f]_\mu \in L^p(\mathcal{E}, \mu)$. Then,

  $$\int_E |f|^p d\mu = 0,$$

  and by the vanishing property for non-negative functions, $|f|^p = 0$, or equivalently, $f = 0$ a.e. $[\mu]$. This indicates that $[f]_\mu = [0_\mathcal{F}]_\mu$, where $[0_\mathcal{F}]_\mu$ is the additive identity on $L^p(\mathcal{E}, \mu)$. Conversely, we can see that

  $$\|[0_\mathcal{F}]_\mu\|_p = \left(\int_E |0_\mathcal{F}|^p d\mu\right)^{\frac{1}{p}} = 0.$$

  Therefore, $\|[f]_\mu\| = 0$ if and only if $[f]_\mu = [0_\mathcal{F}]_\mu$.

- For any $[f]_\mu \in L^p(\mathcal{E}, \mu)$ and $z \in \mathbb{C}$,

  $$\begin{aligned}\|z \cdot [f]_\mu\|_p = \|[zf]_\mu\|_p &= \left(\int_E |zf|^p d\mu\right)^{\frac{1}{p}} \\ &= \left(\int_E |z|^p |f|^p d\mu\right)^{\frac{1}{p}} = |z| \cdot \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}} = |z| \cdot \|[f]_\mu\|_p\end{aligned}$$

  by the linearity of integration.

- For any $[f]_\mu, [g]_\mu \in L^p(\mathcal{E}, \mu)$,

$$\begin{aligned}
\|[f]_\mu + [g]_\mu\|_p = \|[f+g]_\mu\|_p &= \left(\int_E |f+g|^p d\mu\right)^{\frac{1}{p}} \\
&\leq \left(\int_E (|f|+|g|)^p \, d\mu\right)^{\frac{1}{p}} && \text{(The Triangle Inequality)} \\
&\leq \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int_E |g|^p d\mu\right)^{\frac{1}{p}} && \text{(Minkowski's inequality)} \\
&= \|[f]_\mu\|_p + \|[g]_\mu\|_p.
\end{aligned}$$

Therefore, $\|\cdot\|_p$ is a well-defined norm on $L^p(\mathcal{E}, \mu)$, and $(L^P(\mathcal{E}, \mu), \|\cdot\|_p)$ is a normed vector space over the complex field. Since the definition of the L^p norm $\|\cdot\|_p$ uses only the representative f of any equivalence class $[f]_\mu$ in $L^p(\mathcal{E}, \mu)$, we can define

$$\|f\|_p = \|[f]_\mu\|_p = \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}}$$

for any $\mathcal{E}$-measurable complex function f. Note that, for any two such functions f, g, $\|f\|_p = \|g\|_p$ if and only if $[f]_\mu = [g]_\mu$.

Letting $d_p : L^p(\mathcal{E}, \mu) \times L^p(\mathcal{E}, \mu) \to \mathbb{R}_+$ be the metric on $L^p(\mathcal{E}, \mu)$ induced by the L^p norm, defined as

$$d_p([f]_\mu, [g]_\mu) = \|[f]_\mu - [g]_\mu\|_p$$

for any $[f]_\mu, [g]_\mu \in L^p(\mathcal{E}, \mu)$, $(L^p(\mathcal{E}, \mu), d_p)$ is a metric space. If a sequence $\{[f_n]_\mu\}_{n \in N_+}$ in $L^p(\mathcal{E}, \mu)$ converges to some $[f]_\mu \in L^p(\mathcal{E}, \mu)$ in the metric d_p, that is, if

$$\lim_{n\to\infty} d_p([f_n]_\mu, [f]_\mu) = 0,$$

then we say that $\{[f_n]_\mu\}_{n \in N_+}$ converges to $[f]_\mu$ in L^p.
Similarly, for any sequence of complex functions $\{f_n\}_{n \in N_+}$ such that $\int_E |f_n|^p d\mu < +\infty$ for any $n \in N_+$ satisfies

$$\lim_{n\to\infty} \|f_n - f\|_p = 0$$

for some complex valued function f such that $\int_E |f|^p d\mu < +\infty$, we say that $\{f_n\}_{n \in N_+}$ converges to f in L^p, and denote this relation by

$$f_n \xrightarrow{L^p} f.$$

5.3 The Completeness of L^p Spaces

Let $(E, \mathcal{E}, \mu)$ be a measure space and $(L^p(\mathcal{E}, \mu), \|\cdot\|_p)$ the normed vector space defined above, for some $1 \leq p < +\infty$.
In this section, we show that $(L^p(\mathcal{E}, \mu), \|\cdot\|_p)$ is a Banach space over the complex field, that is, the normed vector space $(L^p(\mathcal{E}, \mu), \|\cdot\|_p)$ is a complete metric space with respect to the metric d_p induced by the norm $\|\cdot\|_p$.

Theorem 5.6 (The Riesz-Fischer Theorem)
Let $(E, \mathcal{E}, \mu)$ be a measure space and $\|\cdot\|_p$ the corresponding L^p space for some $1 \leq p < +\infty$. Then, for any sequence $\{f_n\}_{n \in N_+}$ of $\mathcal{E}$-measurable complex valued functions such that $\|f_n\|_p < +\infty$ for any $n \in N_+$ and

$$\lim_{n,m \to \infty} \|f_n - f_m\|_p = 0,$$

there exists a $\mathcal{E}$-measurable complex function f such that $\|f\|_p < +\infty$ and

$$\lim_{n \to \infty} \|f_n - f\|_p = 0.$$

Proof) By assumption, there exists an $n_1 \in N_+$ such that

$$\|f_n - f_m\|_p < 2$$

for any $n, m \geq n_1$.
Now suppose that, for some $k \geq 1$, we have chosen natural numbers $n_1 < \cdots < n_k$ such that, for $1 \leq i \leq k$, $\|f_n - f_m\|_p < 2^{-i}$ for any $n, m \geq n_i$. Then, we can choose an $n_{k+1} \in N_+$ such that $n_{k+1} > n_k$ and

$$\|f_n - f_m\|_p < 2^{-(k+1)}$$

for any $n, m \geq n_{k+1}$.
Constructing the subsequence $\{f_{n_i}\}_{i \in N_+}$ in the above manner, we can see that, for any $i \in N_+$,

$$\|f_{n_{i+1}} - f_{n_i}\|_p < 2^{-i}$$

because $n_{i+1}, n_i \geq n_i$.

Define the sequence $\{g_k\}_{k \in N_+}$ of $\mathcal{E}$-measurable non-negative functions as

$$g_k = \sum_{i=1}^{k} |f_{n_{i+1}} - f_{n_i}|$$

for any $k \in N_+$. Because all the summands are non-negative,

$$g = \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}| = \lim_{k \to \infty} g_k$$

is a well-defined $\mathcal{E}$-measurable non-negative function. It now follows that

$$g_k^p \to g^p$$

pointwise as $k \to \infty$. By Fatou's lemma and Minkowski's inequality,

$$\begin{aligned}
\|g\|_p &= \left(\int_E g^p d\mu\right)^{\frac{1}{p}} = \left(\int_E \left(\liminf_{k\to\infty} g_k^p\right) d\mu\right)^{\frac{1}{p}} \\
&\leq \left(\liminf_{k\to\infty} \int_E g_k^p d\mu\right)^{\frac{1}{p}} && \text{(Fatou's lemma)} \\
&= \liminf_{k\to\infty} \|g_k\|_p && (x \mapsto x^{\frac{1}{p}} \text{ is a continuous mapping}) \\
&\leq \sum_{i=1}^{\infty} \|f_{n_{i+1}} - f_{n_i}\|_p && \text{(Minkowski's inequality and non-negative summands)} \\
&\leq \sum_{i=1}^{\infty} 2^{-i} = 1. && (\text{Choice of } n_1 < n_2 < \cdots)
\end{aligned}$$

Therefore, g^p is μ-integrable, which tells us by the finiteness property that $g^p < +\infty$, or equivalently, $g < +\infty$, a.e. $[\mu]$. In other words, the series

$$\sum_{i=1}^{\infty} (f_{n_{i+1}} - f_{n_i})$$

converges absolutely for μ-almoset every $x \in E$.

Let $E_0 \in \mathcal{E}$ be the almost sure set on which the above series converges absolutely, and define

$$f_x = \sum_{i=1}^{\infty} (f_{n_{i+1}}(x) - f_{n_i}(x)) + f_{n_1}(x) \in \mathbb{C}$$

for any $x \in E_0$. Let the function $f : E \to \mathbb{C}$ be defined as

$$f(x) = \begin{cases} f_x & \text{if } x \in E_0 \\ 0 & \text{otherwise} \end{cases}.$$

Then, for any $x \in E$,

$$f(x) = \lim_{k\to\infty} \left(\sum_{i=1}^{k} (f_{n_{i+1}}(x) - f_{n_i}(x)) + f_{n_1}(x)\right) = \lim_{k\to\infty} f_{n_{k+1}}(x),$$

so that the sequence $\{f_{n_i}\}_{i \in N_+}$ converges pointwise to f for μ-almost every $x \in E$.

Since f can be viewed as the pointwise limit of the sequence of measurable functions $\{f_{n_i} \cdot I_{E_0}\}_{i \in N_+}$, f is itself a $\mathcal{E}$-measurable complex function.

It now remains to show that f is the L^p limit of the sequence $\{f_n\}_{n \in N_+}$. For any $\varepsilon > 0$, by assumption there exists an $N \in N_+$ such that

$$\|f_n - f_m\|_p < \varepsilon$$

for any $n, m \geq N$. Since $\{n_i\}_{i \in N_+}$ is a subsequence of N_+,

$$\|f_{n_i} - f_m\|_p < \varepsilon$$

for any $m \geq N$ and large enough i, which implies that

$$\liminf_{i \to \infty} \|f_{n_i} - f_m\|_p < \varepsilon$$

for any $m \geq N$.
As such, for any $m \geq N$,

$$\begin{aligned} \|f - f_m\|_p = \left(\int_E |f - f_m|^p d\mu \right)^{\frac{1}{p}} &\leq \liminf_{i \to \infty} \left(\int_E |f_{n_i} \cdot I_{E_0} - f_m|^p d\mu \right)^{\frac{1}{p}} && \text{(Fatou's lemma)} \\ &= \liminf_{i \to \infty} \left(\int_E |f_{n_i} - f_m|^p d\mu \right)^{\frac{1}{p}} && (\mu(E_0^c) = 0) \\ &= \liminf_{i \to \infty} \|f_{n_i} - f_m\|_p < \varepsilon. \end{aligned}$$

This holds for any $\varepsilon > 0$, so

$$\lim_{m \to \infty} \|f - f_m\|_p = 0,$$

and because this implies that there exists an $N \in N_+$ such that

$$\|f - f_m\|_p < 1$$

for any $m \geq N$, we have

$$\|f\|_p \leq \|f - f_N\|_p + \|f_N\|_p < +\infty$$

by Minkowski's inequality. This tells us that f is a $\mathcal{E}$-measurable complex function such that $\|f\|_p < +\infty$ and

$$\lim_{n \to \infty} \|f_n - f\|_p = 0.$$

Q.E.D.

In the course of proving the above theorem, we also showed the following corollary:

Corollary to the Riesz-Fischer Theorem Letting $(E, \mathcal{E}, \mu)$ and $\{f_n\}_{n \in N_+}$ satisfy the hypotheses of the Riesz-Fischer theorem, there exists a $\mathcal{E}$-measurable complex function f and a subsequence $\{f_{n_i}\}_{i \in N_+}$ of $\{f_n\}_{n \in N_+}$ such that

$$\lim_{i \to \infty} f_{n_i}(x) = f(x)$$

for μ-almost every $x \in E$.

Proof) This result was proved during the proof of the above theorem.
Q.E.D.

From the Riesz-Fischer theorem we can easily deduce that $(L^p(\mathcal{E}, \mu), d_p)$ is a complete metric space. Let $\{[f_n]_\mu\}_{n \in N_+} \subset L^p(\mathcal{E}, \mu)$ be a sequence that is Cauchy in d_p, that is,

$$\lim_{n,m \to \infty} d_p([f_n]_\mu, [f_m]_\mu) = \lim_{n,m \to \infty} \|f_n - f_m\|_p = 0.$$

This tells us that $\{f_n\}_{n \in N_+}$ is a sequence of $\mathcal{E}$-measurable complex functions that satisfies the conditions of the Riesz-Fischer theorem, and as such that there exists a $\mathcal{E}$-measurable complex function f such that $\|f\|_p < +\infty$ and

$$\lim_{n \to \infty} \|f_n - f\|_p = 0.$$

This makes $[f]_\mu \in L^p(\mathcal{E}, \mu)$ is the L^p limit of the sequence $\{[f_n]_\mu\}_{n \in N_+}$, and we have shown that any sequence in $L^p(\mathcal{E}, \mu)$ that is Cauchy in d_p is also convergent in d_p.

5.4 L^∞ Spaces

5.4.1 The Essential Supremum

Let $(E, \mathcal{E}, \mu)$ be a measure space, and consider a $\mathcal{E}$-measurable non-negative function f. Define the set $S \subset \mathbb{R}$ as

$$S = \{\alpha \in \mathbb{R} \mid \mu(f^{-1}((\alpha, +\infty])) = 0\},$$

that is, the collection of all real numbers α such that f is less than or equal to α a.e. $[\mu]$. Now define $\beta \in [0, +\infty]$ as

$$\beta = \inf S,$$

where we adopt the convention that the infimum of an empty set is $+\infty$. If $S \neq \emptyset$, then $\beta \in \mathbb{R}_+$ by the least upper bound property of the real line, since S is nonempty in this case and bounded below by 0.
Heuristically, β is the upper bound of f on sets that are not of measure zero. For this reason, β is called the essential supremum of f.
Note that, if $S \neq \emptyset$, β must be an element of S, since

$$f^{-1}((\beta, +\infty]) = \bigcup_n f^{-1}\left(\left(\beta + \frac{1}{n}, +\infty\right]\right)$$

and thus

$$\mu(f^{-1}((\beta, +\infty])) \leq \sum_{n=1}^{\infty} \mu\left(f^{-1}\left(\left(\beta + \frac{1}{n}, +\infty\right]\right)\right) = 0,$$

by countable additivity; each summand on the right hand side is 0 because, by the definition of the infimum, for any $n \in N_+$ there exists an $\alpha \in S$ such that $\beta \leq \alpha < \beta + \frac{1}{n}$, so that

$$\mu\left(f^{-1}\left(\left(\beta + \frac{1}{n}, +\infty\right]\right)\right) \leq \mu(f^{-1}((\alpha, +\infty])) = 0.$$

For any $\mathcal{E}$-measurable function f, we define $\|f\|_\infty$ as the essential supremun of $|f| \in \mathcal{E}_+$. The following are properties of the essential supremum:

Lemma 5.7 Let $(E, \mathcal{E}, \mu)$ be a measure space, f, g complex-valued $\mathcal{E}$-measurable functions, and $z \in \mathbb{C}$. Then, the following inequalities hold:

i) $\|0_{\mathcal{F}}\|_\infty = 0$; the eseential supremum of the 0 function is 0.

ii) $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$; the essential supremum satisfies the triangle inequality.

iii) $\|zf\|_\infty = |z| \cdot \|f\|_\infty$; the essential supremum is invariant under scalar multiplication.

iv) If $f = g$ a.e. $[\mu]$, then $\|f\|_\infty = \|g\|_\infty$; the essential supremum is invariant for functions that are almost everywhere equivalent.

Proof) i) Note that

$$\{0_{\mathcal{F}} > \alpha\} = \begin{cases} \emptyset & \text{if } \alpha \geq 0 \\ E & \text{if } \alpha < 0 \end{cases},$$

which implies that

$$\{\alpha \in \mathbb{R} \mid \mu(0_{\mathcal{F}}^{-1}((\alpha, +\infty])) = 0\} = [0, +\infty).$$

Therefore,

$$\|0_{\mathcal{F}}\|_\infty = \inf[0, +\infty) = 0.$$

ii) For any $\mathcal{E}$-measurable function f and g, if $\|f\|_\infty = +\infty$ or $\|g\|_\infty = +\infty$, then the inequality

$$\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$$

holds trivially, so assume that $\|f\|_\infty, \|g\|_\infty \in [0, +\infty)$, or in other words, that the sets

$$S_f = \{\alpha \in \mathbb{R} \mid \mu(|f|^{-1}((\alpha, +\infty])) = 0\}$$

and

$$S_g = \{\alpha \in \mathbb{R} \mid \mu(|g|^{-1}((\alpha, +\infty])) = 0\}$$

are nonempty.
In this case, defining

$$S_{f+g} = \{\alpha \in \mathbb{R} \mid \mu(|f+g|^{-1}((\alpha, +\infty])) = 0\},$$

because $|f+g| \leq |f|+|g|$ by the triangle inequality, for any $\alpha \in S_f$ and $\beta \in S_g$,

$$|f+g|^{-1}((\alpha+\beta,+\infty]) \subset |f|^{-1}((\alpha,+\infty]) \cup |g|^{-1}((\beta,+\infty]),$$

so that

$$\mu(|f+g|^{-1}((\alpha+\beta,+\infty])) \leq \mu(|f|^{-1}((\alpha,+\infty])) + \mu(|g|^{-1}((\alpha,+\infty])) = 0$$

and $\alpha+\beta \in S_{f+g}$. By implication,

$$\|f+g\|_\infty = \inf S_{f+g} \leq \alpha+\beta$$

by the definition of the infimum, and because this holds for any $\alpha \in S_f$ and $\beta \in S_g$,

$$\|f+g\|_\infty \leq \inf S_f + \inf S_g = \|f\|_\infty + \|g\|_\infty.$$

iii) If $z=0$, then $zf=0$ on E and

$$\|zf\|_\infty = \|0_{\mathcal{F}}\|_\infty = 0 = |z| \cdot \|f\|_\infty.$$

Now suppose $z \neq 0$, and define

$$S_{zf} = \{\alpha \in \mathbb{R} \mid \mu(|zf|^{-1}((\alpha,+\infty])) = 0\}$$

and

$$S_f = \{\alpha \in \mathbb{R} \mid \mu(|f|^{-1}((\alpha,+\infty])) = 0\}.$$

For any $\alpha \in S_{zf}$,

$$|zf|^{-1}((\alpha,+\infty]) = \{|zf| > \alpha\} = \left\{|f| > \frac{\alpha}{|z|}\right\} = |f|^{-1}\left(\left(\frac{\alpha}{|z|},+\infty\right]\right),$$

so that

$$\mu\left(|f|^{-1}\left(\left(\frac{\alpha}{|z|},+\infty\right]\right)\right) = \mu(|zf|^{-1}((\alpha,+\infty])) = 0,$$

which implies that $\frac{\alpha}{|z|} \in S_f$. By implication,

$$\|f\|_\infty = \inf S_f \leq \frac{\alpha}{|z|},$$

or $|z| \cdot \|f\|_\infty \leq \alpha$. and because this holds for any $\alpha \in S_{zf}$,

$$|z| \cdot \|f\|_\infty \leq \inf S_{zf} = \|zf\|_\infty.$$

This inequality also implies that

$$\frac{1}{|z|} \cdot \|zf\|_\infty = \left|\frac{1}{z}\right| \cdot \|zf\|_\infty \leq \|f\|_\infty,$$

and as such that

$$\|zf\|_\infty \leq |z| \cdot \|f\|_\infty.$$

Putting the two inequalities together, we have

$$\|zf\|_\infty = |z| \cdot \|f\|_\infty.$$

iv) Assume that $f = g$ a.e. $[\mu]$. Suppose

$$\mu(|f|^{-1}((\alpha, +\infty]) = 0.$$

This implies that

$$0 = \int_E \left(I_{(\alpha,+\infty]} \circ |f|\right) d\mu = \int_E \left(I_{(\alpha,+\infty]} \circ |g|\right) d\mu = \mu(|g|^{-1}((\alpha, +\infty]),$$

where the second equality follows because

$$I_{(\alpha,+\infty]} \circ |f| = I_{(\alpha,+\infty]} \circ |g|$$

a.e. $[\mu]$. It follows that the essential supremum of $|f|$ and $|g|$ are equal, or equivalently, that $\|f\|_\infty = \|g\|_\infty$.

Q.E.D.

5.4.2 L^∞ Spaces as Banach Spaces

Let $(E, \mathcal{E}, \mu)$ be a measure space. By the result proven above, for any $[f]_\mu \in \mathcal{E}/\mathbb{C}$ and $g \in [f]_\mu$,

$$\|f\|_\infty = \|g\|_\infty.$$

Therefore, it makes sense to define

$$\|[f]_\mu\|_\infty = \|f\|_\infty$$

for any equivalence class $[f]_\mu \in \mathcal{E}/\mathbb{C}$ and a representative f of that class.
We can now define $L^\infty(\mathcal{E}, \mu)$ as the following collection of equivalence classes:

$$L^\infty(\mathcal{E}, \mu) = \{[f]_\mu \in \mathcal{E}/\mathbb{C} \mid \|f\|_\infty < +\infty\},$$

that is, as the collection of all equivalence classes whose essential supremum is finite. $L^\infty(\mathcal{E}, \mu)$ is called the collection of all essentially bounded measurable equivalence classes.

Like with L^p spaces for $p \in [1, +\infty)$, $L^\infty(\mathcal{E}, \mu)$ is also a vector space over the complex field. To see this, we first need to verify that it is closed under addition and scalar multiplication. For any $[f]_\mu, [g]_\mu \in L^\infty(\mathcal{E}, \mu)$ and $z \in \mathbb{C}$,

$$\begin{aligned} \|z \cdot [f]_\mu + [g]_\mu\|_\infty &= \|zf + g\|_\infty \\ &\leq \|zf\|_\infty + \|g\|_\infty = |z| \cdot \|f\|_\infty + \|g\|_\infty < +\infty, \end{aligned}$$

so $z \cdot [f]_\mu + [g]_\mu \in L^\infty(\mathcal{E}, \mu)$.
We can show that $L^\infty(\mathcal{E}, \mu)$ satisfies the rest of the vector space axioms through the exact same steps we went through to show that $L^p(\mathcal{E}, \mu)$ is a vector space, so $L^\infty(\mathcal{E}, \mu)$ is a vector space over the complex field.

Furthermore, $(L^\infty(\mathcal{E}, \mu), \|\cdot\|_\infty)$ is a normed vector space. To see this, note that

- If $\|[f]_\mu\|_\infty = 0$, then defining

  $$S = \{\alpha \in \mathbb{R} \mid \mu(|f|^{-1}((\alpha, +\infty])) = 0\},$$

  $\|f\|_\infty = \inf S = 0$. Thus, for any $n \in N_+$, there exists an $\alpha \in S$ such that $\alpha < \frac{1}{n}$; because

  $$|f|^{-1}((1/n, +\infty]) \subset |f|^{-1}((\alpha, +\infty])$$

  and $\mu(|f|^{-1}((\alpha, +\infty])) = 0$, we have $\mu(|f|^{-1}((1/n, +\infty])) = 0$, or equivalently,

  $$\mu\left(\left\{|f| > \frac{1}{n}\right\}\right) = 0.$$

Since

$$\{|f|>0\}=\bigcup_n \left\{|f|>\frac{1}{n}\right\},$$

it follows from countable subadditivity that

$$\mu(\{|f|>0\})\leq \sum_{n=1}^{\infty}\mu\left(\left\{|f|>\frac{1}{n}\right\}\right)=0,$$

so that

$$\mu(\{f\neq 0\})=0.$$

In other words, $f=0_{\mathcal{F}}$ a.e. $[\mu]$, and we have $[f]_\mu=[0_{\mathcal{F}}]_\mu$.
On the other hand,

$$\|[0_{\mathcal{F}}]_\mu\|_\infty=\|0_{\mathcal{F}}\|_\infty=0.$$

We have shown that $\|[f]_\mu\|_\infty=0$ if and only if $[f]_\mu=[0_{\mathcal{F}}]_\mu$.

- For any $z\in\mathbb{C}$ and $[f]_\mu\in L^\infty(\mathcal{E},\mu)$,

$$\|z\cdot[f]_\mu\|_\infty=\|zf\|_\infty=|z|\cdot\|f\|_\infty=|z|\cdot\|[f]_\mu\|_\infty.$$

- For any $[f]_\mu,[g]_\mu\in L^\infty(\mathcal{E},\mu)$,

$$\|[f]_\mu+[g]_\mu\|_\infty=\|f+g\|_\infty\leq\|f\|_\infty+\|g\|_\infty=\|[f]_\mu\|_\infty+\|[g]_\mu\|_\infty.$$

By definition, $\|\cdot\|_\infty$ is a norm on the vector space $L^\infty(\mathcal{E},\mu)$; it is called the L^∞ norm. Letting d_∞ be the metric induced by this norm, $(L^\infty(\mathcal{E},\mu),d_\infty)$ is a metric space, and we can show, like with the L^p-spaces studied earlier, that this metric space is complete.
The following result proves that $(L^\infty(\mathcal{E},\mu),\|\cdot\|_\infty)$ is a Banach space over the complex field.

Theorem 5.8 Let $(E,\mathcal{E},\mu)$ be a measure space, and $\{f_n\}_{n\in N_+}$ a sequence of $\mathcal{E}$-measurable complex functions such that $\|f_n\|_\infty<+\infty$ for any $n\in N_+$ and

$$\lim_{n,m\to\infty}\|f_n-f_m\|_\infty=0.$$

Then, there exists a $\mathcal{E}$-measurable complex function f such that $\|f\|_\infty<+\infty$ and

$$\lim_{n\to\infty}\|f_n-f\|_\infty=0.$$

Proof) For any $k, n, m \in N_+$, by the definition of the essential supremum and the result proven earlier,

$$\mu(\{|f_k| > \|f_k\|_\infty\}) = \mu(\{|f_m - f_n| > \|f_m - f_n\|_\infty\}) = 0.$$

Defining

$$A_k = \{|f_k| > \|f_k\|_\infty\} \quad \text{and} \quad B_{m,n} = \{|f_m - f_n| > \|f_m - f_n\|_\infty\}$$

and

$$E_0 = \bigcup_{k,m,n} (A_k \cup B_{m,n}),$$

it follows from countable suabdditivity that $\mu(E_0) = 0$. Thus,

$$|f_n - f_m| \leq \|f_m - f_n\|_\infty \quad \text{and} \quad |f_k| \leq \|f_k\|_\infty$$

almost everywhere on E, and because $\|f_m - f_n\|_\infty \to 0$ as $m, n \to \infty$,

$$\lim_{n,m\to\infty} |f_n - f_m| = 0$$

uniformly on E_0^c. For any $x \in E_0^c$, this means that $\{f_n(x)\}_{n \in N_+}$ is a Cauchy sequence on the complex plane, and by the completeness of $\mathbb{C}$ under the euclidean metric, there exists an $f_x \in \mathbb{C}$ such that $f_n(x) \to f_x$ as $n \to \infty$. Defining $f : E \to \mathbb{C}$ as

$$f(x) = \begin{cases} f_x & \text{if } x \in E_0^c \\ 0 & \text{otherwise} \end{cases},$$

we can say that $f_n \to f$ pointwise almost everywhere on E, and because $f = \lim_{n\to\infty} f_n \cdot I_{E_0^c}$, f is a $\mathcal{E}$-measurable complex function.

We can easily show that this convergence implies the convergence of $\|f_n - f\|_\infty$ to 0. Choose any $\varepsilon > 0$, and note that, by assumption, there exists an $N \in N_+$ such that

$$\|f_n - f_m\|_\infty < \varepsilon$$

for any $n, m \geq N$. Since $|f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty < \varepsilon$ for any $n, m \geq N$ and $x \in E_0^c$, taking $m \to \infty$ on both sides yields

$$|f_n(x) - f(x)| \leq \varepsilon$$

for any $n \geq N$ and $x \in E_0^c$. In other words,

$$|f_n - f|^{-1}((\varepsilon, +\infty]) \subset E_0$$

for any $n \geq N$, and as such

$$\mu(|f_n - f|^{-1}((\varepsilon, +\infty])) = 0$$

for any $n \geq N$. By implication, for any $n \geq N$,

$$\|f_n - f\|_\infty \leq \varepsilon,$$

and because this holds for any $\varepsilon > 0$, by definition

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0.$$

It remains to show that $\|f\|_\infty < +\infty$, but this follows by noting that there exists an $N \in N_+$ such that

$$\|f_N - f\|_\infty < 1,$$

which is implied by the convergence result above, and

$$\|f\|_\infty \leq \|f_N - f\|_\infty + \|f_N\|_\infty = 1 + \|f_N\|_\infty < +\infty$$

by the triangle inequality.
Q.E.D.

Again, we state the following corollary separately from the theorem above:

Corollary to Theorem 5.8 Letting $(E, \mathcal{E}, \mu)$ and $\{f_n\}_{n\in N_+}$ satisfy the hypotheses of theorem 5.8, there exists a $\mathcal{E}$-measurable complex function f such that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for μ-almost every $x \in E$.

Proof) This result was proved during the proof of the above theorem.
Q.E.D.

From theorem 5.8 we can easily deduce that $(L^\infty(\mathcal{E}, \mu), d_\infty)$ is a complete metric space. Let

$\{[f_n]_\mu\}_{n\in N_+} \subset L^\infty(\mathcal{E},\mu)$ be a sequence that is Cauchy in d_∞, that is,

$$\lim_{n,m\to\infty} \|f_n - f_m\|_\infty = 0.$$

This tells us that $\{f_n\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable complex functions satisfying the conditions of theorem 5.8, and as such that there exists a $\mathcal{E}$-measurable complex function f such that $\|f\|_\infty < +\infty$ and

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0.$$

This makes $[f]_\mu \in L^\infty(\mathcal{E},\mu)$ is the L^∞ limit of the sequence $\{[f_n]_\mu\}_{n\in N_+}$, and we have shown that any sequence in $L^\infty(\mathcal{E},\mu)$ that is Cauchy in d_∞ is also convergent in d_∞.

We have thus shown that L^p spaces for $1 \le p < +\infty$ and L^∞ spaces share much of the same properties: most notably, under their chosen norms, they are Banach spaces over the complex field. For this reason, going forward we use the term L^p space to refer to both spaces with $1 \le p < +\infty$ and those with $p = +\infty$.

5.5 Continuity Properties of L^p Functions

Let (E, τ) be a locally compact Hausdorff space with corresponding Borel σ-algebra $\mathcal{B}(E, \tau)$, and $(E, \mathcal{E}, \mu)$ a measure space satisfying the conditions of the Riesz Representation theorem: to state them for completeness,

i) $\mu(K) < +\infty$ for any compact $K \subset E$

ii) Any $\mathcal{E}$-measurable set is outer regular; for any $A \in \mathcal{E}$,

$$\mu(A) = \inf\{\mu(V) \mid A \subset V, V \in \tau\}$$

iii) Any $\mathcal{E}$-measurable set with finite measure or any open set is inner regular; for any $A \in \mathcal{E}$ such that $\mu(A) < +\infty$ or $A \in \tau$,

$$\mu(A) = \sup\{\mu(K) \mid K \subset A, K \text{ is compact}\}$$

iv) $(E, \mathcal{E}, \mu)$ is complete, and $\mathcal{E}$ contains all Borel sets.

In section 2.4, we proved the continuity of $\mathcal{E}$-measurable complex functions, namely that for any $\mathcal{E}$-measurable complex function f that takes non-zero values on a set of finite measure, there exists a continuously compactly supported complex function that is arbitrarily close to f; this is the content of Lusin's theorem.
Here, we show a similar result for L^p functions, or complex measurable functions f such that $\|f\|_p < +\infty$.

Theorem 5.9 Let (E,τ), $\mathcal{E}$ and μ be the topological space, σ-algeba and measure with the properties stated above. Then, for any $1 \le p < +\infty$, $\varepsilon > 0$ and $\mathcal{E}$-measurable complex function f such that $\|f\|_p < +\infty$, there exists a $g \in C_c(E,\tau)$ such that

$$\|f-g\|_p < \varepsilon.$$

Proof) First note that, for any $g \in C_c(E,\tau)$, $\|g\|_p < +\infty$. To see this, let $K = \overline{\{g \neq 0\}}$; by the extreme value theorem, the continuity of $|g|$ and the compactness of K ensures that there exists some $x^* \in K$ such that $|g(x^*)| = \max_{x \in K} |g(x)|$. Since $g(x) = 0$ for any $x \in K^c$, it follows that

$$\begin{aligned}
\int_E |g|^p d\mu &= \int_K |g|^p d\mu + \int_{K^c} |g|^p d\mu \\
&= \int_K |g|^p d\mu \qquad\qquad (|g(x)| = 0 \text{ for any } x \in K^c) \\
&\le |g(x^*)|^p \cdot \mu(K) < +\infty,
\end{aligned}$$

since $g(x^*) \in \mathbb{C}$ and $\mu(K) < +\infty$ due to the compactness of K. Therefore,

$$\|g\|_p = \left(\int_E |g|^p d\mu\right)^{\frac{1}{p}} < +\infty.$$

To prove the actual claim of the theorem, we first start with indicator functions and build ourselves up to arbitrary complex measurable functions.

Let $A \in \mathcal{E}$ such that $\mu(A) < +\infty$; then,

$$\|I_A\|_p = \left(\int_E I_A d\mu\right)^{\frac{1}{p}} = \mu(A)^{\frac{1}{p}} < +\infty.$$

Because A has finite measure under μ, A is both inner and outer regular. In other words, there exist an open set $V \in \tau$ and a compact set K such that $K \subset A \subset V$ and

$$\mu(V) < \mu(A) + \frac{\varepsilon^p}{2} \quad \text{and} \quad \mu(A) - \frac{\varepsilon^p}{2} < \mu(K).$$

By implication,

$$\begin{aligned}
&\mu(V \setminus A) \le \mu(V) - \mu(A) < \frac{\varepsilon^p}{2}, \\
&\mu(A \setminus K) \le \mu(A) - \mu(K) < \frac{\varepsilon^p}{2}, \quad \text{and} \\
&\mu(V \setminus K) = \mu(V \setminus A) - \mu(A \setminus K) < \varepsilon^p.
\end{aligned}$$

By Urysohn's lemma, there exists a $g \in C_c(E,\tau)$ such that $K \prec g \prec V$, so that $g(x) = 1$

for any $x \in K$, $\overline{\{g \neq 0\}} \subset V$, and $g(x) \in [0,1]$ for any $x \in E$. It follows that $g(x) = I_A(x) = 0$ for $x \in V^c$ and $g(x) = I_A(x) = 1$ for $x \in K$, so that $\{g \neq I_A\} \subset V \setminus K$ and thus

$$|g - I_A| \leq I_{V \setminus K}.$$

Therefore,

$$\begin{aligned}\|g - I_A\|_p &= \left(\int_E |g - I_A|^p d\mu\right)^{\frac{1}{p}} \\ &\leq \left(\int_E I_{V \setminus K} d\mu\right)^{\frac{1}{p}} = \mu(V \setminus K)^{\frac{1}{p}} < \varepsilon.\end{aligned}$$

Let f be a $\mathcal{E}$-measurable simple function such that $\|f\|_p < +\infty$. Letting the canonical form of f be

$$f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i},$$

where $\alpha_1, \cdots, \alpha_n \in (0, +\infty)$ are distinct and $A_1, \cdots, A_n \in \mathcal{E}$ are disjoint,

$$\|f\|_p = \left(\int_E |f|^p d\mu\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n} \alpha_i^p \cdot \mu(A_i)\right)^{\frac{1}{p}} < +\infty,$$

so $\mu(A_1), \cdots, \mu(A_n) < +\infty$. As such, for any $1 \leq i \leq n$, by the previous result there exist $g_1, \cdots, g_n \in C_c(E, \tau)$ such that

$$\|I_{A_i} - g_i\|_p < \frac{\varepsilon}{n \cdot \alpha_i}.$$

Define

$$g = \sum_{i=1}^{n} \alpha_i \cdot g_i;$$

because the linear combination of continuous functions is continuous, g is a continuous non-negative valued function, and since

$$\overline{\{g \neq 0\}} \subset \bigcup_{i=1}^{n} \overline{\{g_i \neq 0\}},$$

where each $\overline{\{g_i \neq 0\}}$ is compact because g_i is compactly supported, $\overline{\{g \neq 0\}}$ is also compact and thus $g \in C_c(E, \tau)$.

Finally, due to the fact that

$$|f - g| \leq \sum_{i=1}^{n} \alpha_i \cdot |I_{A_i} - g_i|,$$

by Minkowski's inequality we have

$$\begin{aligned}\|f-g\|_p &\leq \sum_{i=1}^{n} \alpha_i \cdot \|I_{A_i} - g_i\|_p \\ &< \sum_{i=1}^{n} \frac{\varepsilon}{n} = \varepsilon.\end{aligned}$$

Now let f be a $\mathcal{E}$-measurable non-negative function such that $\|f\|_p < +\infty$. Then, letting $\{s_n\}_{n\in N_+}$ be a sequence of $\mathcal{E}$-measurable simple functions increasing to f, because $s_n \leq f$ and thus $\|s_n\|_p \leq \|f\|_p < +\infty$ for any $n \in N_+$, by the preceding result there exists a $g_n \in C_c(E,\tau)$ such that

$$\|s_n - g_n\|_p < \frac{\varepsilon}{2}.$$

Note that $\{|f-s_n|^p\}_{n\in N_+}$ is a sequence of $\mathcal{E}$-measurable non-negative functions that converges pointwise to 0, and

$$|f-s_n|^p = (f-s_n)^p \leq (f-s_1)^p$$

for any $n \in N_+$, where

$$\|f-s_1\|_p \leq \||f|+|s_1|\|_p \leq \|f_1\|_p + \|s_1\|_p < +\infty$$

by Minkowski's inequality and thus $\int_E |f-s_1|^p d\mu < +\infty$. By the DCT,

$$\begin{aligned}\lim_{n\to\infty} \|f-s_n\|_p &= \left(\lim_{n\to\infty} \int_E |f-s_n|^p d\mu\right)^{\frac{1}{p}} \\ &= \left(\int_E \left(\lim_{n\to\infty} |f-s_n|^p\right) d\mu\right)^{\frac{1}{p}} = 0.\end{aligned}$$

As such, there exist an $N \in N_+$ such that

$$\|f-s_n\|_p < \frac{\varepsilon}{2}$$

for any $n \geq N$; thus, by Minkowski's inequality again,

$$\|f-g_N\|_p = \||f-s_N| + |s_N - g_N|\|_p \leq \|f-s_N\|_p + \|s_N - g_N\|_p < \varepsilon,$$

where $g_N \in C_c(E,\tau)$.

Finally, let f be an arbitrary $\mathcal{E}$-measurable complex function such that $\|f\|_p < +\infty$.

Then,

$$f = \Big(Re(f)^+ - Re(f)^-\Big) + i \cdot \Big(Im(f)^+ - Im(f)^-\Big),$$

where $Re(f)^\pm, Im(f)^\pm \in \mathcal{E}_+$, and because

$$|Re(f)^\pm|^2, |Im(f)^\pm|^2 \le |f|^2 = \Big(Re(f)^+ + Re(f)^-\Big)^2 + \Big(Im(f)^+ + Im(f)^-\Big)^2,$$

$\|Re(f)^\pm\|_p, \|Im(f)^\pm\|_p < +\infty$. From the preceding result, it follows that there exist functions $g_1, g_2 \in C_c(E,\tau)$ and $h_1, h_2 \in C_c(E,\tau)$ such that

$$\Big\|Re(f)^+ - g_1\Big\|_p, \|Re(f)^- - g_2\|_p, \Big\|Im(f)^+ - h_1\Big\|_p, \|Im(f)^- - h_2\|_p < \frac{\varepsilon}{4}.$$

Defiing $g = (g_1 - g_2) + i \cdot (h_1 - h_2) \in C_c(E,\tau)$, we can now see that

$$|f - g| \le \Big|Re(f)^+ - g_1\Big| + |Re(f)^- - g_2| + \Big|Im(f)^+ - h_1\Big| + |Im(f)^- - h_2|,$$

so by Minkowski's inequality,

$$\|f - g\|_p \le \Big\|Re(f)^+ - g_1\Big\|_p + \|Re(f)^- - g_2\|_p + \Big\|Im(f)^+ - h_1\Big\|_p + \|Im(f)^- - h_2\|_p < \varepsilon.$$

Q.E.D.

We conclude with another theorem, which approximates L^p-functions not with continuous functions but this time with functions with finite ranges.

Theorem 5.10 Let $(E,\mathcal{E},\mu)$ be a measure space. Then, for any $1 \le p < +\infty$, $\varepsilon > 0$ and $\mathcal{E}$-measurable function $f : E \to \mathbb{C}$ such that

$$\int_E |f|^p d\mu < +\infty,$$

there exists a $\mathcal{E}$-measurable function s with a finite range satisfying $\mu(\{s \neq 0\}) < +\infty$ such that

$$\|f - s\|_p < \varepsilon.$$

Proof) Suppose f is a non-negative $\mathcal{E}$-measurable function such that $\int_E |f|^p d\mu < +\infty$. Then, there exists a sequence $\{s_n\}_{n\in N_+}$ of $\mathcal{E}$-measurable simple functions increasing to f. By definition, for any $n \in N_+$ we have

$$|f - s_n|^p = (f - s_n)^p \le f^p,$$

where $|f - s_n|^p \to 0$ pointwise as $n \to \infty$. By the DCT,

$$\lim_{n\to\infty} |f - s_n|_p = \left(\lim_{n\to\infty} \int_E |f - s_n|^p d\mu\right)^{\frac{1}{p}} = 0.$$

As such, there exists an $N \in N_+$ such that

$$|f - s_N|_p < \varepsilon.$$

s_N has a finite range (it is a simple function), and by the monotonicity of integration,

$$\int_E s_N^p d\mu \le \int_E f^p d\mu < +\infty,$$

making it an L^p function. Letting the canonical form of s_N be

$$s_N = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$$

for $\alpha_1, \cdots, \alpha_n \in (0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{E}$, because

$$\int_E s_N^p d\mu = \sum_{i=1}^{n} \alpha_i^p \cdot \mu(A_i) < +\infty,$$

$\mu(A_1), \cdots, \mu(A_n) < +\infty$, which implies that

$$\mu(\{s_N \neq 0\}) = \mu(A_1 \cup \cdots \cup A_n) \le \sum_{i=1}^{n} \mu(A_i) < +\infty.$$

Now let f be a complex-valued $\mathcal{E}$-measurable function such that $\int_E |f|^p d\mu < +\infty$. By implication, $Re(f)^{\pm}$ and $Im(f)^{\pm}$ are all $\mathcal{E}$-measurable non-negative L^p-functions, so by the preceding result, there exist $s_+, s_-, h_+, h_- \in \mathcal{E}_+$ with a finite range in $\mathbb{R}_+$ such that $\{s_i \neq 0\}, \{h_i \neq 0\}$ have finite measure under μ for $i = +, -$ and

$$\left\|Re(f)^{\pm} - s_{\pm}\right\|_p < \frac{\varepsilon}{4} \quad \text{and} \quad \left\|Im(f)^{\pm} - h_{\pm}\right\|_p < \frac{\varepsilon}{4}.$$

Defining $g = (s_+ - s_-) + i \cdot (h_+ - h_-)$, g has a finite range,

$$\mu(\{g \neq 0\}) \leq \mu(\{s_+ \neq 0\}) + \mu(\{s_- \neq 0\}) + \mu(\{h_+ \neq 0\}) + \mu(\{h_- \neq 0\}) < +\infty,$$

and

$$\|f - g\|_p \leq \left\|Re(f)^+ - s_+\right\|_p + \left\|Re(f)^- - s_-\right\|_p + \left\|Im(f)^+ - h_+\right\|_p + \left\|Im(f)^- - h_-\right\|_p < \varepsilon$$

by Minkowski's inequality.
Q.E.D.

Another way to phrase the result of the above theorem is to say that the set of all $\mathcal{E}$-measurable complex functions s with finite range such that $\mu(\{s \neq 0\}) < +\infty$ is dense in the set of all L^p functions with respect to the metric d_p.

Chapter 6

Hilbert Space Theory

6.1 Inner Product Spaces

Let H be a complex vector space with additive identity 0_H. The operator $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ is an inner product on H if it satisfies the following conditions:

i) **Conjugate Symmetry**
For any $x, y \in H$,

$$\langle x, y \rangle = \overline{\langle y, x \rangle}$$

ii) **Linearity in First Argument**
For any $x, y, z \in H$ and $\alpha \in \mathbb{C}$,

$$\langle \alpha \cdot x + y, z \rangle = \alpha \cdot \langle x, z \rangle + \langle y, z \rangle.$$

iii) **Positive Definiteness**
For any $x \in H$ such that $x \neq 0_H$,

$$\langle x, x \rangle > 0.$$

The pair $(H, \langle \cdot, \cdot \rangle)$ is called an inner product space over the complex field. The following properties of the inner product follow immediately:

- **Antilinearity in Second Argument**
For any $x, y, z \in H$ and $\alpha \in \mathbb{C}$,

$$\begin{aligned} \langle z, \alpha \cdot x + y \rangle = \overline{\langle \alpha \cdot x + y, z \rangle} &= \overline{\alpha \cdot \langle x, z \rangle + \langle y, z \rangle} \\ &= \overline{\alpha} \cdot \overline{\langle x, z \rangle} + \overline{\langle y, z \rangle} \\ &= \overline{\alpha} \cdot \langle z, x \rangle + \langle z, y \rangle. \end{aligned}$$

- **Inner Product of the Additive Identity**
  For any $x \in H$,

$$\langle 0_H, x\rangle = \langle x - x, x\rangle = \langle x, x\rangle - \langle x, x\rangle = 0.$$

  It follows that

$$\langle x, 0_H\rangle = \overline{\langle 0_H, x\rangle} = 0,$$

  and that $\langle 0_H, 0_H\rangle = 0$. We can thus see that $\langle x, x\rangle = 0$ if and only if $x = 0_H$.

- **Expansion of Inner Product of Sum**
  For any $x, y \in H$,

$$\begin{aligned}
\langle x+y, x+y\rangle &= \langle x, x+y\rangle + \langle y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \langle x, x\rangle + \langle y, y\rangle + \langle x, y\rangle + \overline{\langle x, y\rangle} \\
&= \langle x, x\rangle + 2 \cdot Re(\langle x, y\rangle) + \langle y, y\rangle.
\end{aligned}$$

Given an inner product $\langle \cdot, \cdot \rangle$ on H, we can define the operator $\|\cdot\| : H \to \mathbb{R}_+$ as

$$\|x\| = \sqrt{\langle x, x\rangle}$$

for any $x \in H$.
The following are two useful inequalities involving inner products and $\|\cdot\|$:

Lemma 6.1 Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space over the complex field. The following inequalities hold true:

i) **Cauchy-Schwarz Inequality**
For any $x, y \in H$,

$$|\langle x,y\rangle| \leq \|x\| \cdot \|y\|.$$

ii) **The Triangle Inequality**
For any $x, y \in H$,

$$\|x+y\| \leq \|x\| + \|y\|.$$

iii) **The Parallelogram Law**
For any $x, y \in H$,

$$\|x+y\|^2 + \|x-y\|^2 = 2 \cdot \|x\|^2 + 2 \cdot \|y\|^2.$$

Proof) We first prove the Cauchy-Schwarz inequality. Let $x, y \in H$. If $\|y\| = \sqrt{\langle y,y\rangle} = 0$, then $y = 0_H$ and $\langle x,y\rangle = 0$, so the inequality follows trivially. As such, we assume that $\|y\| > 0$. Define $\alpha = \langle x,y\rangle$. Then, for any $r \in \mathbb{R}$,

$$\begin{aligned} 0 \leq \langle x - r\alpha y, x - r\alpha y\rangle &= \langle x, x - r\alpha y\rangle - r\alpha \cdot \langle y, x - r\alpha y\rangle \\ &= \langle x,x\rangle - r\overline{\alpha} \cdot \langle x,y\rangle - r\alpha \cdot \overline{\langle x,y\rangle} + r^2|\alpha|^2 \cdot \langle y,y\rangle \\ &= \|x\|^2 - 2r \cdot |\langle x,y\rangle|^2 + r^2|\langle x,y\rangle|^2 \cdot \|y\|^2. \end{aligned}$$

Putting $r = \frac{1}{\|y\|^2}$, it follows that

$$0 \leq \|x\|^2 - 2\frac{|\langle x,y\rangle|^2}{\|y\|^2} + \frac{|\langle x,y\rangle|^2}{\|y\|^2},$$

and as such,

$$\frac{|\langle x,y\rangle|^2}{\|y\|^2} \leq \|x\|^2,$$

which implies that $|\langle x,y\rangle|^2 \leq \|x\|^2 \cdot \|y\|^2$. Taking square roots on both sides implies the Cauchy-Schwarz inequality.

The triangle inequality is now easily proven. For any $x, y \in H$,

$$\begin{aligned}
\|x+y\|^2 = \langle x+y, x+y\rangle &= \|x\|^2 + \|y\|^2 + 2\cdot Re(\langle x,y\rangle) \\
&\leq \|x\|^2 + \|y\|^2 + 2\cdot |\langle x,y\rangle| \\
&\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \qquad \text{(Cauchy-Schwarz Inequality)} \\
&= (\|x\| + \|y\|)^2,
\end{aligned}$$

which implies that

$$\|x+y\| \leq \|x\| + \|y\|.$$

Finally, to see that the parallelogram law holds, simply note that

$$\begin{aligned}
\|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\
&= \langle x,x\rangle + 2\cdot Re(\langle x,y\rangle) + \langle y,y\rangle + \langle x,x\rangle + 2\cdot Re(\langle x,-y\rangle) + \langle -y,-y\rangle \\
&= 2\langle x,x\rangle + 2\langle y,y\rangle = 2\|x\|^2 + 2\|y\|^2.
\end{aligned}$$

Q.E.D.

We can now show that $\|\cdot\|$ is a norm on H:

- If $\|x\| = 0$ for some $x \in H$, then $\langle x, x\rangle = 0$ and $x = 0_H$. Conversely, if $x = 0_H$, then $\|x\| = \sqrt{\langle 0_H, 0_H\rangle} = 0$. Thus, $\|x\| = 0$ if and only if $x = 0_H$.

- For any $\alpha \in \mathbb{C}$ and $x \in H$,

$$\begin{aligned}\|\alpha \cdot x\| &= \sqrt{\langle \alpha \cdot x, \alpha \cdot x\rangle} \\ &= \sqrt{\alpha\overline{\alpha} \cdot \langle x, x\rangle} \\ &= |\alpha| \cdot \sqrt{\langle x, x\rangle} = |\alpha| \cdot \|x\|.\end{aligned}$$

- For any $x, y \in H$,

$$\|x + y\| \leq \|x\| + \|y\|$$

  by the above lemma.

We call $\|\cdot\|$ the norm induced by the inner product $\langle \cdot, \cdot\rangle$. Furthermore, letting $d : H \times H \to \mathbb{R}_+$ be defined as

$$d(x, y) = \|x - y\|$$

for any $x, y \in H$, we saw in the previous chapter that d defines a metric on H. We call d the metric induced by the norm $\|\cdot\|$, or the inner product $\langle \cdot, \cdot\rangle$.

Let $(H, \langle \cdot, \cdot\rangle)$ be an inner product space over the complex field and d the metric induced by $\langle \cdot, \cdot\rangle$. The following result shows that we can construct elementary continuous functions on the metric space (H, d) using the inner product:

Lemma 6.2 Let $(H, \langle \cdot, \cdot\rangle)$ be an inner product space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle \cdot, \cdot\rangle$. For any $z \in H$, the mappings

$$x \mapsto \|x\|, \quad x \mapsto \langle x, z\rangle, \quad x \mapsto \langle z, x\rangle$$

defined on E are uniformly continuous with respect to d and the euclidean metric on their target spaces.

Proof) For any $z \in H$, define the functions $f : E \to \mathbb{R}$, $g, h : E \to \mathbb{C}$ as

$$f(x) = \|x\|, \quad g(x) = \langle x, z\rangle \quad \text{and} \quad h(x) = \langle z, x\rangle$$

for any $x \in E$. Because

$$|f(x) - f(y)| = |\|x\| - \|y\|| \leq \|x - y\|$$

for any $x, y \in E$ by the triangle inequality, f is Lipschitz continuous on E.
Likewise, for any $x, y \in E$,

$$|g(x) - g(y)| = |\langle x, z\rangle - \langle y, z\rangle| = |\langle x - y, z\rangle| \leq \|x - y\| \cdot \|z\|$$

and

$$|h(x) - h(y)| = |\langle z, x\rangle - \langle z, y\rangle| = |\langle z, x - y\rangle| \leq \|x - y\| \cdot \|z\|$$

by the (anti)linearity of the inner product and the Cauchy-Schwarz inequality, which tells us that both g and h are Lipschitz continuous.
Q.E.D.

Let $\|\cdot\|$ and d be as above. A subspace V of H is called a closed subspace of H if V is a linear subspace of H that is also closed with respect to the metric d.
Note that, for any linear subspace V of H, the closure $\overline{V}$ of V with respect to the metric d is a closed linear subspace of H. To see this, let V be a subspace of H and note that:

- $0_H \in V \subset \overline{V}$ and

- For any $x, y \in \overline{V}$ and $\alpha \in \mathbb{C}$, there exists sequences $\{x_n\}_{n \in N_+}, \{y_n\}_{n \in N_+} \subset V$ that converge to x, y in the metric d; for any $n \in N_+$,

  $$\|\alpha \cdot x_n + y_n - (\alpha \cdot x + y)\| \leq |\alpha| \cdot \|x_n - x\| + \|y_n - y\|$$

  by the definition of the norm, which tells us that

  $$\lim_{n \to \infty} (\alpha \cdot x_n + y_n) = \alpha \cdot x + y,$$

  and because each $\alpha \cdot x_n + y_n \in V$ by the definitio of a linear subspace, $\alpha \cdot x + y$ is a limit point of V and therefore contained in $\overline{V}$.

6.2 Orthogonality and Finite-Dimensional Subspaces

Through the norm of a vector, normed vector spaces introduce the concept of the magnitude of a vector. Inner product spaces go one step further and allow us to define the angle between two vectors as their inner product. In particular, it is of great interest when two vectors are orthogonal, or when their inner product equals 0. We say two vectors $x,y \in H$ are orthonormal if $\|x\| = \|y\| = 1$ and $\langle x,y\rangle = 0$.

6.2.1 Orthogonal Sets and Orthogonal Complements

We first define the concept of orthogonal sets and introduce orthogonal complements of subspaces.

Let $(H,\langle\cdot,\cdot\rangle)$ be a complex inner product space. A subset V of H is said to be orthogonal if, for any $x,y \in V$ such that $x \neq y$, $\langle x,y\rangle = 0$, that is, if any two distinct elements of V are orthogonal. Note that V is linearly independent if V is an orthogonal set of non-zero vectors; to see this, choose any finite subset $F = \{x_1,\cdots,x_n\}$ of V and suppose

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0_H$$

for some $\alpha_1,\cdots,\alpha_n \in \mathbb{C}$. Then, for any $1 \leq j \leq n$,

$$\begin{aligned} 0 = \langle 0_H, x_j\rangle &= \langle \sum_{i=1}^{n} \alpha_i x_i, x_j\rangle \\ &= \sum_{i=1}^{n} \alpha_i \cdot \langle x_i, x_j\rangle = \alpha_j \cdot \langle x_j, x_j\rangle. \end{aligned}$$

Since $x_j \neq 0_H$, $\langle x_j,x_j\rangle \neq 0$, which implies that $\alpha_j = 0$. This holds for $1 \leq j \leq n$, so $\alpha_1 = \cdots = \alpha_n = 0$ and as such F is linearly independent. This again holds for any finite subset of V, so by definition V is also linearly independent.

For any subspace V of H, the orthogonal complement of V is defined as the collection of all vectors in H that are orthogonal to every element of V: formally,

$$V^{\perp} = \{x \in H \mid \langle x,z\rangle = 0 \text{ for any } z \in V\}.$$

The following are some notable properties of the orthogonal complement of a linear subspace:

Lemma 6.3 Let $(H,\langle\cdot,\cdot\rangle)$ be an inner product space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle\cdot,\cdot\rangle$. For any linear subspace V of H and its orthogonal complement, the following hold true:

i) $V^{\perp}$ is a closed linear subspace of H.

ii) V and $V^{\perp}$ are independent linear subspaces, that is, $x + y = 0_H$ for some $x \in V$ and $y \in V^{\perp}$

implies $x = y = 0_H$.

Proof) i) $V^\perp$ satisfies the following conditions to be a linear subspace:

- $0_H \in V^\perp$ because $\langle 0_H, z\rangle = 0$ for any $z \in V$

- For any $\alpha \in \mathbb{C}$ and $x, y \in V^\perp$, because

$$\langle x, z\rangle = \langle y, z\rangle = 0$$

for any $z \in V$, by the linearity of the inner product we have

$$\langle \alpha \cdot x + y, z\rangle = \alpha \cdot \langle x, z\rangle + \langle y, z\rangle = 0,$$

so that $\alpha \cdot x + y \in V^\perp$ as well.

Finally, we can see that $V^\perp$ is a closed subset of H; note that

$$V^\perp = \bigcap_{x \in V} \{x\}^\perp.$$

For any $x \in V$, define the function $h : H \to \mathbb{C}$ as $h(y) = \langle y, x\rangle$ for any $y \in E$; we saw in lemma 6.2 that h is continuous with respect to the metric d, and because $\{0\}$ is a closed set and

$$\{x\}^\perp = \{y \in H \mid \langle y, x\rangle = 0\} = h^{-1}(\{0\}),$$

$\{x\}^\perp$ is a closed subset of H. As such, $V^\perp$, being the intersection of closed sets, must also be closed.

ii) For any $x \in V$ and $y \in V^\perp$, if $x + y = 0_H$, then

$$0 = \langle 0_H, x\rangle = \langle x + y, x\rangle = \langle x, x\rangle$$

because $\langle y, x\rangle = 0$, so that $x = 0_H$ and thus $y = 0_H$.

Q.E.D.

6.2.2 Finite-Dimensional Orthonormal Sets

A subset V of H is orthonormal if it is an orthogonal set such that $\langle x, x\rangle = \|x\|^2 = 1$ for any $x \in V$, that is, if the norm of each element of V has been normalized to 1. Succinctly, V is an orthonormal set if

$$\langle x, y\rangle = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

for any $x, y \in V$.

Let A be an arbitrary index set and $\{u_\alpha \mid \alpha \in A\}$ an orthonormal set contained in H. For any $x \in H$, we can then define

$$\hat{x}(\alpha) = \langle x, u_\alpha\rangle$$

for any $\alpha \in A$. The collection $\{\hat{x}(\alpha) \mid \alpha \in A\}$ is called the set of Fourier coefficients of x with respect to the orthonormal set $\{u_\alpha \mid \alpha \in A\}$.

The following are propreties of finite orthonormal sets:

Lemma 6.4 Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle\cdot,\cdot\rangle$. Let A be an arbitrary index set, $V = \{u_\alpha \mid \alpha \in A\}$ an orthonormal set, and F a finite subset of A. Denote by M_F the span of $\{u_\alpha \mid \alpha \in F\}$. The following hold true:

i) If $\varphi : A \to \mathbb{C}$ is a complex function that equals 0 ouside of F, then the vector $y \in M_F$ defined as

$$y = \sum_{\alpha \in F} \varphi(\alpha) \cdot u_\alpha$$

satisfies $\hat{y}(\alpha) = \varphi(\alpha)$ for any $\alpha \in A$, and

$$\|y\|^2 = \sum_{\alpha \in F} |\varphi(\alpha)|^2.$$

ii) If $x \in H$ and the vector $s_F(x) \in M_F$ is defined as

$$s_F(x) = \sum_{\alpha \in F} \hat{x}(\alpha) \cdot u_\alpha,$$

then for any $y \in M_F$,

$$\langle x - s_F(x), y\rangle = 0 \quad \text{and} \quad \|x - s_F(x)\| \leq \|x - y\|,$$

where the inequality holds as an equality if and only if $y = s_F(x)$. In addition,

$$\sum_{\alpha \in F} |\hat{x}(\alpha)|^2 \leq \|x\|^2.$$

Proof) i) Define $y = \sum_{\alpha \in F} \varphi(\alpha) \cdot u_\alpha$, as in the statement of the lemma. By definition, for any $\alpha \in A$,

$$\begin{aligned}
\hat{y}(\alpha) = \langle y, u_\alpha \rangle &= \langle \sum_{\alpha' \in F} \varphi(\alpha') \cdot u_{\alpha'}, u_\alpha \rangle \\
&= \sum_{\alpha' \in F} \varphi(\alpha') \cdot \langle u_{\alpha'}, u_\alpha \rangle \\
&= \begin{cases} \varphi(\alpha) & \text{if } \alpha \in F \\ 0 & \text{if } \alpha \notin F \end{cases}
\end{aligned}$$

by definition of V as an orthonormal set. Since $\varphi(\alpha) = 0$ if $\alpha \notin F$, it follows that $\hat{y}(\alpha) = \varphi(\alpha)$ for any $\alpha \in A$.
Furthermore,

$$\begin{aligned}
\|y\|^2 &= \langle \sum_{\alpha \in F} \varphi(\alpha) \cdot u_\alpha, \sum_{\alpha' \in F} \varphi(\alpha') \cdot u_{\alpha'} \rangle \\
&= \sum_{\alpha \in F} \sum_{\alpha' \in F} \varphi(\alpha) \overline{\varphi(\alpha')} \cdot \langle u_\alpha, u_{\alpha'} \rangle \\
&= \sum_{\alpha \in F} |\varphi(\alpha)|^2 \cdot \langle u_\alpha, u_\alpha \rangle \\
&= \sum_{\alpha \in F} |\varphi(\alpha)|^2.
\end{aligned}$$

ii) For any $x \in H$, define $s_F(x)$ as in the statement of the lemma.
For any $v \in M_F$, there exist $b_\alpha \in \mathbb{C}$ for any $\alpha \in F$ such that

$$v = \sum_{\alpha \in F} b_\alpha \cdot u_\alpha.$$

Therefore,

$$\begin{aligned}
\langle x - s_F(x), v \rangle = \langle x, v \rangle - \langle s_F(x), v \rangle &= \langle x, \sum_{\alpha \in F} b_\alpha \cdot u_\alpha \rangle - \langle \sum_{\alpha \in F} \hat{x}(\alpha) \cdot u_\alpha, \sum_{\alpha \in F} b_\alpha \cdot u_\alpha \rangle \\
&= \sum_{\alpha \in F} \overline{b_\alpha} \cdot \langle x, u_\alpha \rangle - \sum_{\alpha \in F} \sum_{\alpha' \in F} \hat{x}(\alpha) \overline{b_{\alpha'}} \cdot \langle u_\alpha, u_{\alpha'} \rangle \\
&= \sum_{\alpha \in F} \overline{b_\alpha} \hat{x}(\alpha) - \sum_{\alpha \in F} \hat{x}(\alpha) \overline{b_\alpha} = 0.
\end{aligned}$$

In addition, for any $y \in M_F$, note that

$$\begin{aligned}\|x-y\|^2 = \langle x-y, x-y\rangle &= \langle (x-s_F(x)) + (s_F(x)-y), (x-s_F(x)) + (s_F(x)-y)\rangle \\ &= \|x-s_F(x)\|^2 + \|s_F(x)-y\|^2 + 2\cdot Re(\langle x-s_F(x), s_F(x)-y\rangle).\end{aligned}$$

Because $s_F(x), y \in M_F$, their difference is also in M_F, and as such

$$\langle x-s_F(x), s_F(x)-y\rangle = 0$$

by the preceding result. This implies that

$$\|x-y\|^2 = \|x-s_F(x)\|^2 + \|s_F(x)-y\|^2 \geq \|x-s_F(x)\|^2,$$

where the equality holds if and only if $y = s_F(x)$.
Finally, putting $y = 0_H$ in the above inequality yields

$$\|x\|^2 = \|x-s_F(x)\|^2 + \|s_F(x)\|^2 \geq \|s_F(x)\|^2,$$

where

$$\|s_F(x)\|^2 = \sum_{\alpha\in F} |\hat{x}(\alpha)|^2.$$

Q.E.D.

6.2.3 Orthogonal Projections

$s_F(x)$ defined in the above lemma is the unqiue solution to the problem

$$\min_{y \in M_F} \|x - y\|.$$

In other words, it is the linear combination of the vectors in $\{u_\alpha \mid \alpha \in F\}$ that is closest to, or best approximates, the vector x in terms of the norm induced by the inner product $\langle \cdot, \cdot \rangle$. We call $s_F(x)$ the orthogonal projection of x on the linear subspace M_F.
More generally, given some subset V of an inner product space $(H, \langle \cdot, \cdot \rangle)$, we call $y \in V$ an orthogonal projection of $x \in H$ on V if

$$\|x - y\| = \inf_{z \in V} \|x - z\|.$$

The following are some general results on orthogonal projections:

Theorem 6.5 Let $(H, \langle \cdot, \cdot \rangle)$ be an inner product space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle \cdot, \cdot \rangle$. Let V be a subset of H. The following hold true:

i) Let $x \in H$, and suppose that $y \in V$ is an orthogonal projection of x on V, that is,

$$\|x - y\| = \inf_{z \in V} \|x - z\|.$$

Then, y is the unique orthogonal projection of x on y if V is a convex set.

ii) Let $x \in H$, and suppose that V is a subspace of H. Then, $y \in V$ is the unique orthogonal projection of x on V if and only if $\langle x - y, z \rangle = 0$ for any $z \in V$.

iii) Let V be a subspace of H.
Suppose that, for any $x \in H$, there exists a unique orthogonal projection of x on V. Then, $H = V \bigoplus V^\perp$.
Moreover, denoting the mapping from x to its unique orthogonal projection on V by P, and the mapping from x to $x - Px$ by Q, P, Q are linear transformations from H into V and $V^\perp$, and Qx is the orthogonal projection of x on $V^\perp$.
For any $x \in H$, we have

$$\|x\|^2 = \|Px\|^2 + \|Qx\|^2.$$

Proof) i) Let $x \in H$, and suppose that $y \in V$ is an orthogonal projection of x on V, that is,

$$\|x - y\| = \inf_{z \in V} \|x - z\|.$$

Suppose that V is a convex set, and let $y' \in V$ be another orthogonal projection of x on V. Denoting $\delta = \|x-y\| = \|x-y'\|$, by the parallelogram law,

$$\left\|\frac{1}{2}(y-y')\right\|^2 + \left\|x - \frac{y+y'}{2}\right\|^2 = 2\cdot\left\|\frac{1}{2}(x-y)\right\|^2 + 2\cdot\left\|\frac{1}{2}(x-y')\right\|^2.$$

Multiplying both sides by 4 and noting that $\frac{y+y'}{2} \in V$ because V is convex, we can see that

$$\begin{aligned}\|y-y'\|^2 &= 2\cdot\left(\|x-y\|^2 + \|x-y'\|^2 - 2\left\|x-\frac{y+y'}{2}\right\|^2\right)\\ &\leq 2\cdot\left(2\delta^2 - 2\delta^2\right) = 0,\end{aligned}$$

since $\left\|x - \frac{y+y'}{2}\right\|^2 \geq \delta^2$. Therefore, $\|y-y'\| = 0$ and $y = y'$, making y the unique orthogonal projection of x on V.

ii) Let V be a subspace of H, and for any $x \in H$, suppose that $y \in V$ is the orthogonal projection of x on V (it is unique because V is convex). Then, by definition,

$$\|x-y\| \leq \|x-z\|$$

for any $z \in V$. Choose any $z \in V$; if $z = 0_H$, then $\langle x-y, z\rangle = 0$ trivially. Supose that $z \neq 0_H$. For any $a \in \mathbb{C}$, $y + az \in V$ because V is a subspace of H, and thus

$$\|x-y\| \leq \|x-(y+az)\| = \|(x-y)-az\|$$

by the definition of y as the orthogonal projection of x on V. Then, we have

$$\begin{aligned}\|x-y\|^2 \leq \|(x-y)-az\|^2 &= \langle (x-y)-az, (x-y)-az\rangle\\ &= \|x-y\|^2 + |a|^2\|z\|^2 - a\cdot\langle z, x-y\rangle - \bar{a}\cdot\langle x-y, z\rangle,\end{aligned}$$

so that

$$0 \leq |a|^2\|z\|^2 - a\cdot\langle z, x-y\rangle - \bar{a}\cdot\langle x-y, z\rangle.$$

Putting $a = \frac{\langle x-y,z\rangle}{\|z\|^2} \in \mathbb{C}$, the above inequality becomes

$$0 \leq \frac{|\langle x-y,z\rangle|^2}{\|z\|^2} - 2\cdot\frac{|\langle x-y,z\rangle|^2}{\|z\|^2} = -\frac{|\langle x-y,z\rangle|^2}{\|z\|^2},$$

and multiplying both sides by $-\|z\|^2$, we obtain

$$|\langle x-y,z\rangle|^2 \leq 0.$$

This implies that $|\langle x-y,z\rangle|^2 = 0$, or that $\langle x-y,z\rangle = 0$.

Now suppose that $y \in V$ satisfies $\langle x-y, z\rangle$ for any $z \in V$. Then, for any $z \in V$,

$$\|x-z\|^2 = \langle (x-y)+(y-z), (x-y)+(y-z)\rangle = \|x-y\|^2 + \|y-z\|^2 + 2\cdot Re\left(\langle x-y, y-z\rangle\right).$$

Since $y-z \in V$ (V is a subspace), by assumption we have $\langle x-y, y-z\rangle = 0$, so that

$$\|x-z\|^2 = \|x-y\|^2 + \|y-z\|^2 \geq \|x-y\|^2.$$

This holds for any $z \in V$, so y is an orthogonal projection of x on V, and by the convexity of V, it is the unique orthogonal projection of x on V.

iii) Let V be a subspace of H, and suppose that, for any $x \in H$, there exists a unique orthogonal projection of x on V. Define the mapping $P : H \to V$ so that Px is the unique orthogonal projection of x on V for any $x \in H$. For any $x \in H$, by the second result, we can see that $\langle x - Px, z\rangle = 0$ for any $z \in V$. This means that $x - Px \in V^\perp$, so defining the mapping $Q : H \to V^\perp$ as $Qx = x - Px$ for any $x \in H$,

$$x = Px + Qx,$$

where $Px \in V$ and $Qx \in V^\perp$, for any $x \in H$. This shows us that $H = V \bigoplus V^\perp$, where the sum becomes a direct sum because V and $V^\perp$ are independent.

To see that P and Q are linear, choose any $x, y \in H$, $a \in \mathbb{C}$, and note that

$$\begin{aligned} a\cdot(Px+Qx)+(Py+Qy) &= a\cdot x + y \\ &= P(ax+y)+Q(ax+y) \end{aligned}$$

by the decomposition above. Rearranging terms yields

$$P(ax+y) - a\cdot Px - Py = a\cdot Qx + Qy - Q(ax+y);$$

the left hand side is in V and the right hand side in $V^\perp$, and because $V \cap V^\perp = 0_H$ (if $z \in V\cap V^\perp$, then $\langle z, z\rangle = \|z\|^2 = 0$, or $z = 0_H$), this tells us that

$$P(ax+y) - a\cdot Px - Py = a\cdot Qx + Qy - Q(ax+y) = 0_H.$$

The linearity of P and Q follows immediately.

For any $x \in H$ and $y \in V^\perp$,

$$\langle x - Qx, y\rangle = \langle Px, y\rangle = 0$$

because $Px \in V$; by the previous result, this tells us that $Qx \in V^{\perp}$ is the unique orthogonal projection of x on $V^{\perp}$.

Finally, choose any $x \in H$, and note that

$$\|x\|^2 = \|Px + Qx\|^2 = \|Px\|^2 + \|Qx\|^2 + 2 \cdot Re(\langle Px, Qx \rangle) = \|Px\|^2 + \|Qx\|^2$$

because $Px \in V$ and $Qx \in V^{\perp}$.

Q.E.D.

Note that, for any finite dimensional subspace V of an inner product space $(H, \langle \cdot, \cdot \rangle)$, lemma 6.4 tells us that, for any $x \in H$, there exists a unique orthogonal projection of x on V. This means that result iii) above applies to any finite dimensional subspace V of H: specifically, $H = V \bigoplus V^{\perp}$, and there exist linear mappings P, Q from H into $V, V^{\perp}$ such that Px is the orthogonal projection of x on V, Qx is that of x on $V^{\perp}$, and $x = Px + Qx$ for any $x \in H$.

Of course, we cannot even assert the existence of the orthogonal projection if the subspace in question is not finite-dimensional. It is one of the most important properties of Hilbert spaces that any closed convex subspace has a unique orthogonal projection.

6.3 Properties of Hilbert Spaces

Let $(H, \langle\cdot,\cdot\rangle)$ be an inner product space over the complex field, and $\|\cdot\|$ and d the norm and metric induced by $\langle\cdot,\cdot\rangle$. If the metric space (H,d) is complete, then we call $(H, \langle\cdot,\cdot\rangle)$ a Hilbert space.

6.3.1 The Projection Theorem

If V is a finite-dimensional linear subspace of an inner product space $(H, \langle\cdot,\cdot\rangle)$, we saw above in theorem 6.5 that, for any $x \in H$, there always exists an orthogonal projection of x onto V and as such that H can be expressed as the direct sum of V and its orthogonal complement $V^{\perp}$.

In general, there does not always exist an orthogonal projection of a vector $x \in H$ onto an arbitrary subset V of H. However, Hilbert spaces are special in that, for any closed convex subset V of H and some $x \in H$, there always exists an orthogonal projection of x onto V.
This property, called the Hilbert projection theorem, allows us to work with orthgonal projections without worrying about their existence in infinite-dimensional spaces (for example, function spaces like L^2 spaces), and as such forms the cornerstone of many important mathematical results, including but not limited to the Radon-Nikodym theorem and the characterization of conditional expectations.
The projection theorem is stated below:

Theorem 6.6 (The Hilbert Projection Theorem)
Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle\cdot,\cdot\rangle$. For any nonempty closed convex subset V of H and $x \in H$, there exists a unique $y \in V$ such that

$$\|x - y\| = \inf_{z \in V} \|x - z\|.$$

Furthermore, if V is a closed subspace of H, then the following hold true:

i) $H = V \bigoplus V^{\perp}$.

ii) Defining Px as the unique orthogonal projection of x on V for any $x \in H$, the mapping $x \mapsto Px$ is a linear transformation from H into V.

iii) Defining $Qx = x - Px$ for any $x \in H$, the mapping $x \mapsto Qx$ is a linear transformation from H into $V^{\perp}$, and Qx is an orthogonal projection of x on $V^{\perp}$ for any $x \in H$.

iv) For any $x \in H$, $x = Px + Qx$ and

$$\|x\|^2 = \|Px\|^2 + \|Qx\|^2.$$

Proof) Choose any $x \in H$. Since the set

$$\{z \in V \mid \|x - z\|\}$$

is nonempty due to the nonemptiness of V and bounded below by 0, the infimum

$$\delta = \inf_{z \in V} \|x - z\|$$

exists in $\mathbb{R}_+$. For any $n \in N_+$, by the definition of the infimum there exists a $y_n \in V$ such that

$$\delta \leq \|x - y_n\| < \delta + \frac{1}{n},$$

or equivalently,

$$|\|x - y_n\| - \delta| < \frac{1}{n},$$

so the sequence $\{\|x - y_n\|\}_{n \in N_+}$ converges to δ.
For any $m, n \in N_+$, by the parallelogram law we can see that

$$\left\|\frac{1}{2}(y_n - y_m)\right\|^2 + \left\|x - \frac{y_n + y_m}{2}\right\|^2 = 2 \cdot \left\|\frac{1}{2}(x - y_n)\right\|^2 + 2 \cdot \left\|\frac{1}{2}(x - y_m)\right\|^2,$$

and because $\frac{y_n + y_m}{2} \in V$ by the convexity of V,

$$\delta^2 = \inf_{z \in V} \|x - z\|^2 \leq \left\|x - \frac{y_n + y_m}{2}\right\|^2$$

and we have

$$\|y_n - y_m\|^2 \leq 2\|x - y_n)\|^2 + 2\|x - y_m\|^2 - 4\delta^2.$$

Taking $n, m \to \infty$ on both sides, since

$$\lim_{n \to \infty} \|x - y_n)\|^2 = \lim_{m \to \infty} \|x - y_m\|^2 = \delta^2,$$

the right hand side converges to 0 and thus

$$\lim_{n, m \to \infty} \|y_n - y_m\|^2 = 0.$$

This shows us that $\{y_n\}_{n \in N_+} \subset V$ is Cauchy in the metric d; by the completeness of the metric space (H, d), there exists a $y^* \in H$ such that $y_n \to y^*$ as $n \to \infty$ in the metric d. Finally, because V is a closed subset of H and $\{y_n\}_{n \in N_+}$ is a sequence in V, $y^* \in V$

as well. The continuity of the mapping $y \mapsto \|x-y\|$ on H now tells us that

$$\|x-y^*\| = \lim_{n\to\infty} \|x-y_n\| = \delta = \inf_{z\in V} \|x-z\|.$$

We have shown so far that y^* is an orthogonal projection of x on V. Because V is convex, by theorem 6.5, y^* is the unique orthogonal projection of x on V.

Suppose that V is a closed subspace of H. Then, because V is a closed convex subset of H, by the result above, for any $x \in H$ there exists a unique orthogonal projection of x on V. By theorem 6.5, we can now see that properties i) to iv) above hold true.

Q.E.D.

Corollary to the Hilbert Projection Theorem Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle\cdot,\cdot\rangle$. For any nonempty closed convex subset V of H, there exists a unique $y \in V$ of smallest norm, that is, a unique element $y \in V$ such that $\|y\| \leq \|z\|$ for any $z \in V$.

Proof) This follows immediately from the Hilbert Projection Theorem. Specifically, because V is a closed convex subset of the hilbert space H, there exists a unique $y \in V$ such that

$$\|y\| = \|0_H - y\| = \inf_{z\in V} \|0_H - z\| = \inf_{z\in V} \|z\|.$$

Q.E.D.

6.3.2 A Representation Theorem

A useful application of the Projection Theorem is the Representation Theorem, which tells us, much like the Riesz representation theorem studied in chapter 4, that any linear functional on a Hilbert space can be represented as the inner product with some element of that space. Coincidentally, this theorem is also called the Riesz representation theorem, so in order to distinguish between the two representation theorems, we will call the present theorem the Riesz representation theorem for Hilbert spaces.
The statement and proof of the theorem are given below:

Theorem 6.7 (The Riesz-Fréchet Representation Theorem)
Let $(H, \langle \cdot,\cdot \rangle)$ be a Hilbert space over the complex field and $\|\cdot\|$ and d the norm and metric induced by $\langle \cdot,\cdot \rangle$. For any continuous linear functional $L \in \mathcal{L}(H, \mathbb{C})$, there exists a unique element $\varphi \in H$ (also called the Riesz representation of L) such that

$$L(x) = \langle x, \varphi \rangle$$

for any $x \in H$.

Proof) We first show uniqueness. Suppose that there exist $\varphi_1, \varphi_2 \in H$ such that

$$L(x) = \langle x, \varphi_i \rangle$$

for any $x \in H$ and $i = 1,2$. Then,

$$L(\varphi_1 - \varphi_2) = \langle \varphi_1 - \varphi_2, \varphi_1 \rangle = \langle \varphi_1 - \varphi_2, \varphi_2 \rangle,$$

so that

$$\|\varphi_1 - \varphi_2\|^2 = \langle \varphi_1 - \varphi_2, \varphi_1 \rangle - \langle \varphi_1 - \varphi_2, \varphi_2 \rangle = 0.$$

This implies that $\varphi_1 = \varphi_2$, and that the Riesz representation of L, if it exists, is unique.

To show existence, first define V as the null space of L. Since $V = L^{-1}(\{0\})$, where $\{0\}$ is closed and L continuous with respect to the metric d, V is a closed subset of H. Furthermore, V is a linear subspace, so by the Hilbert Projection Theorem, $H = V \bigoplus V^{\perp}$, that is, for any $x \in H$ there exists a unique $P(x) \in V$ such that $x - P(x) \in V^{\perp}$. If the rank of L is 0, then $L(x) = 0$ for any $x \in H$ and thus $\varphi = 0_H$.
Suppose now that the rank of L is 1 (full rank). In this case, there exists a $x \in H$ such that $L(x) \neq 0$, and V is a proper subset of H. Then, $x - P(x) \in V^{\perp}$ but $x - P(x) \neq 0_H$ because $P(x) \in V$ but $x \notin V$, which tells us that $V^{\perp} \neq \{0_H\}$.

Choose some $z \in V^\perp$ such that $|z| = 1$, and for any $x \in H$, define

$$u(x) = L(x) \cdot z - L(z) \cdot x.$$

It follows that

$$L(u(x)) = L(x) \cdot L(z) - L(z) \cdot L(x) = 0$$

by linearity, so $u(x) \in V$ and $\langle u(x), z \rangle = 0$. Therefore,

$$0 = \langle u(x), z \rangle = \langle L(x) \cdot z - L(z) \cdot x, z \rangle = L(x) \cdot \langle z, z \rangle - L(z) \cdot \langle x, z \rangle = L(x) - \langle x, \overline{L(z)} \cdot z \rangle,$$

and rearranging terms, we have

$$L(x) = \langle x, \overline{L(z)} \cdot z \rangle.$$

This holds for any $x \in H$, so it follows that $\varphi = \overline{L(z)} \cdot z$.
Q.E.D.

6.4 Hilbert Spaces and L^2 Spaces

The most important Hilbert space are L^2 spaces. In this section, we will show both that L^2 spaces are Hilbert spaces, and that any Hilbert space can be represented as an L^2 space.

6.4.1 L^2 Spaces as Hilbert Spaces

Let $(E,\mathcal{E},\mu)$ be a measure space, and $L^2(\mathcal{E},\mu)$ the L^2-space associated with this measure space. Define the operation $\langle\cdot,\cdot\rangle_2 : L^2(\mathcal{E},\mu)\times L^2(\mathcal{E},\mu)\to\mathbb{C}$ as

$$\langle [f]_\mu,[g]_\mu\rangle_2 = \int_E f\overline{g}d\mu$$

for any $[f]_\mu,[g]_\mu\in L^2(\mathcal{E},\mu)$; this integral is well-defined because both f and $\overline{g}$ are μ-integrable. $\langle\cdot,\cdot\rangle_2$ is easily shown to be an inner product on $L^2(\mathcal{E},\mu)$:

- For any $[f]_\mu,[g]_\mu\in L^2(\mathcal{E},\mu)$, becausae $Re(f),Im(f),Re(g),Im(g)$ are all μ-integrable real-valued random variables,

  $$\begin{aligned}\langle [f]_\mu,[g]_\mu\rangle_2 = \int_E f\overline{g}d\mu &= \int_E (Re(f)+i\cdot Im(f))\,(Re(g)-i\cdot Im(g))\,d\mu\\ &= \int_E (Re(f)Re(g)+Im(f)Im(g))\,d\mu + i\cdot\int_E (Im(f)Re(g)-Re(f)Im(g))\,d\mu\\ &= \overline{\int_E (Re(f)Re(g)+Im(f)Im(g))\,d\mu + i\cdot\int_E (Re(f)Im(g)-Im(f)Re(g))\,d\mu}\\ &= \overline{\int_E (Re(f)-i\cdot Im(f))\,(Re(g)+i\cdot Im(g))\,d\mu}\\ &= \overline{\int_E \overline{f}gd\mu} = \overline{\langle [g]_\mu,[f]_\mu\rangle_2}\end{aligned}$$

  by the linearity of integration.

- For any $[f]_\mu,[g]_\mu,[h]_\mu\in L^2(\mathcal{E},\mu)$ and $z\in\mathbb{C}$,

  $$\begin{aligned}\langle z\cdot[f]_\mu+[g]_\mu,[h]_\mu\rangle_2 &= \int_E (zf+g)\overline{h}d\mu\\ &= z\cdot\int_E f\overline{h}d\mu+\int_E g\overline{h}d\mu\\ &= z\cdot\langle [f]_\mu,[h]_\mu\rangle_2+\langle [g]_\mu,[h]_\mu\rangle_2\end{aligned}$$

  again by the linearity of integration.

- For any $[f]_\mu\in L^2(\mathcal{E},\mu)$ such that $[f]_\mu\neq[0_\mathcal{F}]_\mu$,

  $$\langle [f]_\mu,[f]_\mu\rangle_2 = \int_E |f|^2d\mu\geq 0$$

because $|f|^2$ is a non-negative $\mathcal{E}$-measurable function. If $\langle [f]_\mu, [f]_\mu \rangle_2 = 0$, then by the vanishing property of non-negative functions, $|f|^2 = 0$, or equivalently, $f = 0_\mathcal{F}$ a.e. $[\mu]$. This contradicts the assumption that $[f]_\mu \neq [0_\mathcal{F}]_\mu$, so

$$\langle [f]_\mu, [f]_\mu \rangle_2 > 0.$$

Therefore, $(L^2(\mathcal{E},\mu), \langle \cdot, \cdot \rangle_2)$ is an inner product space over the complex field. Furthermore, the norm $\|\cdot\|_2$ induced by $\langle \cdot, \cdot \rangle_2$ is defined as

$$\|[f]_\mu\|_2 = \sqrt{\langle [f]_\mu, [f]_\mu \rangle_2} = \left(\int_E |f|^2 d\mu \right)^{\frac{1}{2}},$$

so $\|\cdot\|_2$ is just the L^2 norm on $L^2(\mathcal{E},\mu)$. The Riesz-Fischer Theorem showed that $(L^2(\mathcal{E},\mu), \|\cdot\|_2)$ is a Banach space, so by extension $(L^2(\mathcal{E},\mu), \langle \cdot, \cdot \rangle_2)$ is a Hilbert space.

Note also that, for the Hilbert space $(L^2(\mathcal{E},\mu), \langle \cdot, \cdot \rangle_2)$, the Cauchy Schwarz inequality and triangle inequality are just versions of Hölder's inequality and Minkowski's inequality.

One of the most important examples of an application of Hilbert space methods to L^2 spaces is the Radon-Nikodym theorem, which will be studied in the next section.

6.4.2 Integrals with respect to the Counting Measure

Before moving onto the converse result, that any Hilbert space can be represented as an L^2 space, we first focus on the representation of integrals with respect to the counting measure on an arbitrary set.

Let A be an arbitrary set and $\mathcal{A} = 2^A$ its discrete σ-algebra. For any non-negative function $\varphi : A \to [0, +\infty]$, which is trivially $\mathcal{A}$-measurable, we define

$$\sum_{\alpha \in A} \varphi(\alpha) = \sup_{F \subset A,\ F \text{ is finite}} \sum_{\alpha \in F} \varphi(\alpha),$$

that is, as the supremum of the sum of φ over finite subsets of A.
The following is the main result of this section:

Lemma 6.8 Let A be an arbitrary set and c the counting measure on $(A, 2^A)$. Then, for any non-negative function $\varphi : A \to [0, +\infty]$,

$$\sum_{\alpha \in A} \varphi(\alpha) = \int_A \varphi dc,$$

that is, the sum of the values of φ on A is exactly that of the integral of φ with respect to the counting measure on A.
Furthermore, if φ is c-integrable, that is, $\sum_{\alpha \in A} \varphi(\alpha) < +\infty$, then

$$\{\alpha \in A \mid \varphi(\alpha) \neq 0\}$$

is at most countable.

Proof) Suppose that $\varphi(\alpha) = +\infty$ for some $\alpha \in A$. Then, $\varphi \cdot I_{\{\alpha\}} \leq \varphi$ on A, and by the monotonicity of integration,

$$\int_A \varphi dc \geq \int_A \left(\varphi \cdot I_{\{\alpha\}}\right) dc = \varphi(\alpha) = +\infty,$$

which implies that $\int_A \varphi dc = +\infty$. Similarly, because $\{\alpha\}$ is a finite subset of A,

$$+\infty = \varphi(\alpha) \leq \sum_{\alpha' \in A} \varphi(\alpha),$$

so that

$$\sum_{\alpha \in A} \varphi(\alpha) = +\infty = \int_A \varphi dc.$$

Now suppose that $\varphi(\alpha) < +\infty$ for any $\alpha \in A$.

This means that, for any finite subset F of A, the function

$$f = \sum_{\alpha \in F} \varphi(\alpha) \cdot I_{\{\alpha\}}$$

is a (trivially) measurable simple function, with integral

$$\int_A f dc = \sum_{\alpha \in F} \varphi(\alpha).$$

Because $f \le \varphi$ on A, by the monotonicity of integration,

$$\sum_{\alpha \in F} \varphi(\alpha) = \int_A f dc \le \int_A \varphi dc.$$

This holds for any finite subset F of A, so

$$\sum_{\alpha \in A} \varphi(\alpha) \le \int_A \varphi dc.$$

To establish the reverse inequality, we may assume that $\sum_{\alpha \in A} \varphi(\alpha) < +\infty$ without loss of generality (otherwise, it is trivial). Suppose that the inequality above is strict. Then, by the definition of $\int_A \varphi dc$ as the supremum of measurable simple functions on A majorized by φ, there exists a simple function s on A such that $s \le \varphi$ on A and

$$\sum_{\alpha \in A} \varphi(\alpha) < \int_A s dc.$$

Letting $\{a_1, \cdots, a_n\} \subset (0, +\infty)$ be the finite non-zero values s takes, for any finite subset $F \subset A$

$$\sum_{\alpha \in F} s(\alpha) \le \sum_{\alpha \in F} \varphi(\alpha) \le \sum_{\alpha \in A} \varphi(\alpha)$$

because $s \le \varphi$ on A. By implication,

$$\sum_{\alpha \in A} s(\alpha) \le \sum_{\alpha \in A} \varphi(\alpha) < +\infty,$$

For any $1 \le i \le n$, this means that $s^{-1}(\{a_i\})$ is a finite set, since if $s^{-1}(\{a_i\})$ contains more than $\frac{1}{a_i} \sum_{\alpha \in A} \varphi(\alpha)$ elements, then choosing a finite subset J of $s^{-1}(\{a_i\})$ with more than $\frac{1}{a_i} \sum_{\alpha \in A} \varphi(\alpha)$ elements, we have

$$\sum_{\alpha \in A} s(\alpha) \ge \sum_{\alpha \in J} s(\alpha) = a_i \cdot |J| > \sum_{\alpha \in A} \varphi(\alpha),$$

a contradiction. Therefore, $\{s \ne 0\} = s^{-1}(\{a_1\}) \cup \cdots \cup s^{-1}(\{a_n\})$ must be a finite set.

Defining $F = \{s \neq 0\}$, we can now see that

$$\int_A s dc = \sum_{\alpha \in F} s(\alpha) \leq \sum_{\alpha \in A} s(\alpha)$$

but

$$\int_A s dc = \sum_{\alpha \in F} s(\alpha) > \sum_{\alpha \in A} \varphi(\alpha),$$

a contradiction. It follows that

$$\sum_{\alpha \in A} \varphi(\alpha) = \int_A \varphi dc.$$

Finally, to show the second result, suppose $\sum_{\alpha \in A} \varphi(\alpha) < +\infty$. This means that there exists a natural number M such that $\sum_{\alpha \in A} \varphi(A) < M$.
For any $n \in N_+$, define the set

$$A_n = \{\alpha \in A \mid \varphi(\alpha) > \frac{1}{n}\}.$$

If A_n contains more than nM elements, then we can choose a finite subset J of A_n with more than nM elements, for which

$$\sum_{\alpha \in A} \varphi(\alpha) \geq \sum_{\alpha \in J} \varphi(\alpha) > \frac{1}{n}|J| > M,$$

since $\varphi(\alpha) > \frac{1}{n}$ for any $\alpha \in J$. This is a contradiction, so A_n must contain at most nM elements; A_n is a finite set.
This holds for any $n \in N_+$, and

$$\{\alpha \in A \mid \varphi(\alpha) \neq 0\} = \bigcup_n \{\alpha \in A \mid \varphi(\alpha) > \frac{1}{n}\} = \bigcup_n A_n,$$

so φ is non-zero for at most countably many points in A.
Q.E.D.

We have thus shown that the integral of a non-negative function with respect to the counting measure is the sum of that function over the set on which it is defined.

For any $1 \leq p \leq +\infty$,the L^p space on A with respect to the counting measure c on A, denoted $L^p(2^A, c)$, is often denoted $\ell^p(A)$.
If two functions $\varphi, \xi : A \to \mathbb{C}$ are equivalent a.e. $[c]$, then by definition

$$c(\{\varphi \neq \xi\}) = 0.$$

This means that $\{\varphi \neq \xi\}$ contains 0 elements, or that it is the empty set, which in turn indicates that $\varphi = \xi$ everywhere on A. In other words, the equivalence class $[\varphi]_c = \{\varphi\}$, so that $\ell^p(A)$ does not need to be understood as the collection of equivalence classes, but rather as the collection of complex functions on A.

If $1 \leq p < +\infty$, $\varphi \in \ell^p(A)$ if and only if

$$\int_A |\varphi|^p dc = \sum_{\alpha \in A} |\varphi(\alpha)|^p < +\infty,$$

so $\ell^p(A)$ can be viewed as the collection of functions φ on A such that

$$\sum_{\alpha \in A} |\varphi(\alpha)|^p < +\infty.$$

For any $\varphi : A \to \mathbb{C}$, defining

$$S = \{\beta \in \mathbb{R} \mid c(|\varphi|^{-1}((\beta, +\infty])) = 0\} = \left[\sup_{\alpha \in A} |\varphi(\alpha)|, +\infty\right],$$

the essential supremum of $|\varphi|$ with respect to the counting measure is

$$\|\varphi\|_\infty = \inf S = \sup_{\alpha \in A} |\varphi(\alpha)|.$$

Therefore, $\ell^\infty(A)$ is the collection of all bounded functions φ on A.

For any $1 \leq p < +\infty$, by theorem 5.10 the set of all complex-valued functions s on A with finite range such that $c(\{s \neq 0\}) < +\infty$ is dense in $\ell^p(A)$. $c(\{s \neq 0\}) < +\infty$ implies that $\{s \neq 0\}$ is a finite subset of A, that is, if $s(\alpha) \neq 0$ for a finite number of $\alpha \in A$. Conversely, if some complex-valued function $s : A \to [0, +\infty]$ is non-zero on a finite subset F of A, then s has finite range $\{0\} \cup \{s(\alpha) \mid \alpha \in F\}$, and $c(\{s \neq 0\}) = c(F) < +\infty$.
Therefore, the collection of all complex-valued functions s on A that are non-zero only for finitely many $\alpha \in A$ is dense in $\ell^p(A)$ for any $1 \leq p < +\infty$.

6.4.3 Aribtrary Orthonormal Sets

So far, we have studied the property of finite orthonormal sets in inner product spaces and the projection property of closed convex subsets of Hilbert spaces. We now extend our analysis to arbtirary orthonormal sets in Hilbert spaces. To this end, we introduce the concept of the isometry.

Let (E,d) and (F,ρ) be metric spaces. We say a function $f : E \to F$ is an isometry from (E,d) to (F,ρ) if

$$d(x,y) = \rho(f(x), f(y))$$

for any $x, y \in E$. In other words, it is a distance-preserving function; in this sense, it is iso-metric. Note the following result concerning isometries:

Lemma 6.9 Let (E,d) and (F,ρ) be metric spaces, where (E,d) is complete. Suppose that $f : E \to F$ is a function such that

i) f is continuous

ii) There exists a subset E_0 of E that is dense in E such that

$$d(x,y) = \rho(f(x), f(y))$$

for any $x, y \in E_0$, and

iii) The image $f(E_0)$ is dense in F.

Then, f is an isometry from (E,d) onto (F,ρ).

Proof) We first show that f is an isometry from (E,d) to (F,ρ).
For any $x, y \in E$ and $\varepsilon > 0$, by the continuity of f on E, there exists a $0 < \delta < \frac{\varepsilon}{2}$ such that

$$\rho(f(x), f(z)) < \frac{\varepsilon}{2} \quad \rho(f(y), f(w)) < \frac{\varepsilon}{2}$$

for any $z, w \in E$ such that $d(x,z), d(y,w) < \delta$.
By the denseness of E_0 in E, there exist $x_0, y_0 \in E_0$ such that $d(x,x_0), d(y,y_0) < \delta$, so that

$$\rho(f(x), f(x_0)) < \frac{\varepsilon}{2} \quad \rho(f(y), f(y_0)) < \frac{\varepsilon}{2}.$$

Since f is isometric on E_0,

$$d(x_0, y_0) = \rho(f(x_0), f(y_0)),$$

and by the triangle inequality,

$$|\rho(f(x), f(y)) - \rho(f(x_0), f(y_0))| < \rho(f(x), f(x_0)) + \rho(f(y), f(y_0)) < \varepsilon$$

and

$$|d(x,y) - d(x_0,y_0)| \leq d(x,x_0) + d(y,y_0) < 2 \cdot \delta < \varepsilon.$$

By implication,

$$|d(x,y) - \rho(f(x), f(y))| \leq |d(x,y) - d(x_0,y_0)| + |\rho(f(x), f(y)) - \rho(f(x_0), f(y_0))| < 2\varepsilon.$$

This holds for any $\varepsilon > 0$, so

$$d(x,y) = \rho(f(x), f(y)),$$

and f is an isometry from (E,d) into (F,ρ).

To show that f maps onto F, choose any $y \in F$. Because $f(E_0)$ is dense in F, there exists a sequence $\{y_n\}_{n \in N_+} \subset f(E_0)$ that converges to y in the metric ρ. For any $n \in N_+$, there exists an $x_n \in N_+$ such that $f(x_n) = y_n$, since $y_n \in f(E_0)$. f is an isometry and $\{y_n\}_{n \in N_+}$, being a convergent sequence, is also Cauchy in ρ, so

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

indicating that $\{x_n\}_{n \in N_+}$ is Cauchy in d. Finally, by the completeness of (E,d), there exists an $x \in E$ such that $x_n \to x$ in the metric d, and by the continuity of f, we can see that

$$f(x) = \lim_{n \to \infty} f(x_n) = \lim_{n \to \infty} y_n = y.$$

This indicates that $y = f(x) \in f(E)$, and as such that f maps onto F.
Q.E.D.

The next theorem uses the fact that, for any Hilbert space $(H, \langle \cdot, \cdot \rangle)$ with norm $\|\cdot\|$ and metric d, (H,d) is complete by the Riesz-Fischer theorem. It follows that, for any closed subset V of H, (V,d) is also a complete metric space.

Theorem 6.10 Let $(H, \langle\cdot,\cdot\rangle)$ be a complex Hilbert space with induced norm and metric $\|\cdot\|$ and d. Suppose A is an arbitrary index set and $V = \{u_\alpha \mid \alpha \in A\}$ an orthonormal subset of H. Denote by P the span of V. Then, for any $x \in H$,

$$\sum_{\alpha\in A} |\hat{x}(\alpha)|^2 \leq \|x\|^2, \qquad \text{(Bessel's Inequality)}$$

and the mapping $x \mapsto \hat{x}$ is a continuous mapping from H into $\ell^2(A)$ whose restriction to the closure $\overline{P}$ of P is an isometry from $\overline{P}$ onto $\ell^2(A)$.

Proof) For any $x \in H$, recall that the function $\hat{x}$ on A is defined as

$$\hat{x}(\alpha) = \langle x, u_\alpha\rangle$$

for any $\alpha \in A$. For any finite subset F of A, since $\{u_\alpha \mid \alpha \in F\}$ is a finite orthonormal set in H, by lemma 6.4 we have

$$\sum_{\alpha\in F} |\hat{x}(\alpha)|^2 \leq \|x\|^2.$$

As such,

$$\sum_{\alpha\in A} |\hat{x}(\alpha)|^2 = \sup_{F\subset A,\ F \text{ is finite}} \sum_{\alpha\in F} |\hat{x}(\alpha)|^2 \leq \|x\|^2.$$

Since $\|x\|^2 < +\infty$, this implies that $\sum_{\alpha\in A} |\hat{x}(\alpha)|^2 < +\infty$, and as such that $\hat{x} \in \ell^2(A)$. Let $\langle\cdot,\cdot\rangle_2$ be the inner product on $\ell^2(A)$ defined as

$$\langle\varphi,\xi\rangle_2 = \int_A \varphi\overline{\xi}dc = \sum_{\alpha\in A} \varphi(\alpha)\overline{\xi}(\alpha)$$

for any $\varphi, \xi \in \ell^2(A)$, and $\|\cdot\|_2$ and d_2 the norm and metric induced by this inner product.

Define the function $f : H \to \ell^2(A)$ as $f(x) = \hat{x}$ for any $x \in H$. For any $a \in \mathbb{C}$ and $x, y \in H$, denote $z = ax + y$ and note that

$$\hat{z}(\alpha) = \langle z, u_\alpha\rangle = a\cdot\langle x, u_\alpha\rangle + \langle y, u_\alpha\rangle = a\cdot\hat{x}(\alpha) + \hat{y}(\alpha)$$

for any $\alpha \in A$, so that

$$f(ax+y) = \hat{z} = a\cdot\hat{x}(\alpha) + \hat{y}(\alpha) = a\cdot f(x) + f(y).$$

We have thus shown that f is a linear transformation mapping from H into $\ell^2(A)$.

Furthermore, for any $x, y \in H$,

$$\begin{aligned}
\|f(x) - f(y)\|_2^2 = \|\hat{x} - \hat{y}\|_2^2 &= \int_A |\hat{x} - \hat{y}|^2 dc \\
&= \sum_{\alpha \in A} |\hat{x}(\alpha) - \hat{y}(\alpha)|^2 = \sum_{\alpha \in A} \left| (x \hat{-} y)(\alpha) \right|^2 && \text{(Linearity of } f\text{)} \\
&\leq \|x - y\|^2. && \text{(Bessel's inequality)}
\end{aligned}$$

Therefore, f is a continuous lienar transformation with respect to the metrics d and d_2.

Finally, let $g : \overline{P} \to \ell^2(A)$ be the restriction of f to $\overline{P}$. We will show that $(\overline{P}, d)$, $(\ell^2(A), d_2)$ and g satisfy the conditions of lemma 6.8.
Because $\overline{P}$ is a closed subset of H, $(\overline{P}, d)$ is a complete metric space. g is continuous on $\overline{P}$ due to the continuity of f on H.
P is a subset of $\overline{P}$ that is dense in $\overline{P}$. Choose any $x, y \in P$; because x, y are finite linear combinations of the vectors in V, there exists a finite subset F of A such that

$$x = \sum_{\alpha \in F} x_\alpha \cdot u_\alpha \quad \text{and} \quad y = \sum_{\alpha \in F} y_\alpha \cdot u_\alpha.$$

It follows that

$$\hat{x}(\alpha) = \begin{cases} x_\alpha & \text{if } \alpha \in F \\ 0 & \text{if } \alpha \notin F \end{cases} \quad \text{and} \quad \hat{y}(\alpha) = \begin{cases} y_\alpha & \text{if } \alpha \in F \\ 0 & \text{if } \alpha \notin F \end{cases},$$

so that

$$\begin{aligned}
d_2(g(x), g(y))^2 = \|\hat{x} - \hat{y}\|_2^2 &= \sum_{\alpha \in A} |\hat{x}(\alpha) - \hat{y}(\alpha)|^2 \\
&= \sum_{\alpha \in F} |x_\alpha - y_\alpha|^2 = \left\| \sum_{\alpha \in F} (x_\alpha - y_\alpha) \cdot u_\alpha \right\|^2 \\
&= \|x - y\|^2 = d(x, y)^2.
\end{aligned}$$

This shows us that g is an isometry on P.
Finally, for any $y \in g(P)$, there exists an $x \in P$ such that $g(x) = \hat{x} = y$. Since x is a finite linear combination of the vectors in V, we saw previously that $\hat{x}$ is a complex-valued function on A that is non-zero for only finitely many $\alpha \in A$, specifically the elements of F.
Conversely, suppose that y is some complex-valued function on A that is non-zero for only finitely many $\alpha \in A$. Letting $y \neq 0$ on the finite set F, define $y_\alpha = y(\alpha)$ for any

$\alpha \in F$. Then, $y = \hat{x} = g(x)$, where $x \in H$ is defined as

$$x = \sum_{\alpha \in F} y_\alpha \cdot u_\alpha.$$

Since x is a finite linear combination of the vectors in V, $x \in P$, and $y \in g(P)$. Therefore, $g(P)$ is precisely the set of all complex-valued functions on A that are non-zero only for finitely many $\alpha \in A$, and as such is dense in $\ell^2(A)$ with respect to the metric d_2.

By lemma 6.9, it now follows that g is an isometry from $(\overline{P}, d)$ onto $(\ell^2(A), d_2)$.
Q.E.D.

Since $\overline{P}$ is a subset of H, the above theorem essentially tells us that any function on A that is square-summable is the Fourier coefficient of some vector in H.
The next result furnishes sufficient conditions for $\overline{P}$ in the previous theorem to coincide with the entire space H.

Theorem 6.11 Let $(H, \langle \cdot, \cdot \rangle)$ be a complex Hilbert space with induced norm and metric $\|\cdot\|$ and d. Suppose A is an arbitrary index set and $V = \{u_\alpha \mid \alpha \in A\}$ an orthonormal subset of H. Denote by P the span of V. Then, the following are equivalent:

i) V is a maximal orthonormal subset of H; that is, there does not exist any $x \in H$ such that $x \notin V$ and $V \cup \{x\}$ is an orthonormal subset of H

ii) The span P of V is dense in H; that is, $\overline{P} = H$

iii) For any $x \in H$,

$$\sum_{\alpha \in A} |\hat{x}(\alpha)|^2 = \|x\|^2$$

iv) **(Parseval's Identity)** For any $x, y \in H$,

$$\sum_{\alpha \in A} \hat{x}(\alpha)\overline{\hat{y}(\alpha)} = \langle x, y \rangle.$$

Proof) We will show that ii) $\rightarrow$ iii) $\rightarrow$ iv) $\rightarrow$ i) $\rightarrow$ ii).

ii) $\rightarrow$ iii)

If ii) holds, then by theorem 6.10, the mapping $x \mapsto \hat{x}$ is a continuous isometry from H onto $\ell^2(A)$, so that, for any $x, y \in H$,

$$\sum_{\alpha \in A} |\hat{x}(\alpha) - \hat{y}(\alpha)|^2 = \int_A |\hat{x} - \hat{y}|^2 dc = \|\hat{x} - \hat{y}\|_2^2 = \|x - y\|^2.$$

Putting $y = 0_H$ leads to the equality iii).

iii) $\rightarrow$ iv)

We make use of the polar identity: for any $x, y \in H$,

$$\begin{aligned}\|x+y\|^2 - \|x-y\|^2 + i \cdot \|x+iy\|^2 - i \cdot \|x-iy\|^2 &= 4 \cdot Re(\langle x, y \rangle) + 4i \cdot Im(\langle x, y \rangle) \\ &= 4\left(Re(\langle x, y \rangle) + i \cdot Im(\langle x, y \rangle)\right) = 4 \cdot \langle x, y \rangle.\end{aligned}$$

Therefore, using iii) and the polar identity again, we have

$$\begin{aligned}4\cdot\langle x,y\rangle &= \|\hat{x}+\hat{y}\|_2^2-\|\hat{x}-\hat{y}\|_2^2+i\cdot\|\hat{x}+i\hat{y}\|_2^2-i\cdot\|\hat{x}-i\hat{y}\|_2^2\\ &= 4\cdot\langle\hat{x},\hat{y}\rangle_2 = 4\cdot\int_A \hat{x}\overline{\hat{y}}dc = 4\cdot\sum_{\alpha\in A}\hat{x}(\alpha)\overline{\hat{y}(\alpha)},\end{aligned}$$

which implies iv).

iv) $\to$ i)

Suppose that iv) is true but i) is not. Since this means V is not the maximally orthonormal subset of H, there exists some $u\in H$ that is not contained in V but $\{u\}\cup V$ is an orthonormal subset of H. By definition, $\|u\|=1$, so $u\neq 0_H$, and $\langle u,u_\alpha\rangle=0$ for any $\alpha\in A$. Putting $x=y=u$, by Parseval's identity,

$$\begin{aligned}\sum_{\alpha\in A}|\hat{u}(\alpha)|^2 &= \sum_{\alpha\in A}\hat{x}(\alpha)\overline{\hat{y}(\alpha)}\\ &= \langle x,y\rangle = \|u\|^2 = 1.\end{aligned}$$

However, because

$$\hat{u}(\alpha)=\langle u,u_\alpha\rangle=0$$

for any $\alpha\in A$, $\sum_{\alpha\in A}|\hat{u}(\alpha)|^2=0$, which implies the equality $0=1$, a contradiction. Therefore, i) must hold true.

i) $\to$ ii)

Finally, suppose that i) holds, but ii) does not. $\overline{P}$ is a closed subset of H, and choosing any $x,y\in\overline{P}$, there exist sequences $\{x_n\}_{n\in N_+}$ and $\{y_n\}_{n\in N_+}$ in P that converge to x,y. For any $a\in\mathbb{C}$, because P is a subpsace of H, $a\cdot x_n+y_n\in P$ for any $n\in N_+$, and the limit $a\cdot x+y$ of the sequence $\{a\cdot x_n+y_n\}_{n\in N_+}$ must then belong in $\overline{P}$. Finally, because $0_H\in P$, 0_H belongs in $\overline{P}$ as well. Therefore, $\overline{P}$ is a closed subspace of H, and by the Hilbert projection theorem, $H=\overline{P}\bigoplus\overline{P}^{\perp}$.

We have assumed that $\overline{P}\neq H$, that is, that $\overline{P}$ is a proper subset of H. This indicates that there exists an $x\in H$ such that $x\neq\overline{P}$, so this x can be decomposed into the sum of $x_1\in\overline{P}$ and $x_2\in\overline{P}^{\perp}$, where $x_2\neq 0_H$ (otherwise, $x=x_1\in\overline{P}$, a contradiction). Thus, $\overline{P}^{\perp}$ contains a non-zero element.

Choose $u\in\overline{P}^{\perp}$ so that $\|u\|=1$, which exists because $\overline{P}^{\perp}$ contains a non-zero element and $\overline{P}^{\perp}$ is a subspace of H. Defining $V'=\{u\}\cup V$, we can see that, for any $\alpha\in A$, because $V \subset P \subset \overline{P}$,

$$\langle u, u_\alpha \rangle = 0.$$

This indicates that V' is an orthonormal subset of H, which contradicts the maximality of V. Therefore, ii) must hold, and $\overline{P} = H$.

Q.E.D.

A maximally orthonormal subset of an inner product space $(H, \langle \cdot, \cdot \rangle)$ is called an orthonormal basis, or a complete orthonormal subset, of H.

6.4.4 Hilbert Spaces as L^2 Spaces

Suppose that we know a maximally orthonormal set exists for any complex Hilbert space $(H, \langle\cdot,\cdot\rangle)$. In this case, letting $V = \{u_\alpha \mid \alpha \in A\}$ be the maximal orthonormal set in question, by the theorem above the span of V is dense in H, and by the theorem before that, we can tell that the mapping $x \mapsto \hat{x}$ from H into $\ell^2(A)$ is a linear transformation and an isometry from H onto $\ell^2(A)$.

This means that the mapping $x \mapsto \hat{x}$ becomes a surjective linear transformation from H into $\ell^2(A)$. The isometry also tells us that, for any $x, y \in H$, if $\hat{x} = \hat{y}$, then

$$0 = \|\hat{x} - \hat{y}\|_2 = \|x - y\|$$

and that $x = y$, so that the mapping $x \mapsto \hat{x}$ is a bijective linear transformation. Denote this mapping by $L : H \to \ell^2(A)$.

Finally, for any $a \in \mathbb{C}$ and $\varphi, \xi \in \ell^2(A)$, letting $x, y \in H$ be chosen so that $L(x) = \hat{x} = \varphi$ and $L(y) = \hat{y} = \xi$, we have

$$\begin{aligned} L^{-1}(a\varphi + \xi) = L^{-1}(a \cdot L(x) + L(y)) &= L^{-1}(L(ax + y)) \\ &= ax + y = a \cdot L^{-1}(\varphi) + L^{-1}(\xi). \end{aligned}$$

In other words, the inverse mapping L^{-1} is also a linear transformation, which implies that L is an isomorphism from H into $\ell^2(A)$.

Therefore, if we can just establish that H has a maximally orthonormal set, then the Hilbert space H is isomorphic to the L^2-space $\ell^2(A)$, and the isomorphism in question is exactly the mapping that maps each vector of H to its Fourier coefficient. It is in this sense that any Hilbert space can be regarded as an L^2-space.

It is easy to establish that any Hilbert space has a maximally orthonormal subset with help from the Hausdorff Maximality Theorem. Recall from the beginning of this text that, given that the axiom of choice holds, so does the Hausdorff Maximality Theorem, which posits that any partially ordered set contains a maximal totally ordered set, or a totally ordered subset that does not have any totally ordered expansion. The main theorem is given below:

Theorem 6.12 Let $(H, \langle\cdot,\cdot\rangle)$ be a complex inner product space. Every orthonormal set V is contained in a maximally orthonormal set W, that is, an orthonormal set such that, for any $x \notin W$, $W \cup \{x\}$ is not orthonormal.

Proof) Let $\mathcal{F}$ be the collection of all orthonormal sets in H that contain V. $\mathcal{F}$ is a nonempty partially ordered set under the set inclusion operator $\subset$, so by the Hausdorff maximality theorem, it has a subset $\mathcal{G}$ that is maximally totally ordered under set inclusion. Let W be the union of every element of $\mathcal{G}$. We will show that W is our desired maximal orthonormal set.

Choose any $x, y \in W$. Since $W = \bigcup_{B \in \mathcal{G}} B$, there exist $A, B \in \mathcal{G}$ such that $x \in A$ and $y \in B$. $\mathcal{G}$ is totally ordered under set inclusion, so either $A \subset B$ or $B \subset A$; assuming $A \subset B$ without loss of generality, $x, y \in B$, and because B is an orthonormal set containing V by definition, $\langle x, y \rangle = 1$ if $x = y$ and $\langle x, y \rangle = 0$ otherwise. This shows us that W is an orthonormal set.

To show that W is maximal, choose some $x \notin W$ and define $W^* = W \cup \{x\}$. If W^* is an orthonormal set, then, because it contains V through W, $W^* \in \mathcal{F}$. W is a proper subset of W^*, so $W^* \notin \mathcal{G}$. The collection $\mathcal{G}^* = \mathcal{G} \cup \{W^*\}$ is then totally ordered under set inclusion since, for any $A \in \mathcal{G}$, $A \subset W \subset W^*$, and $\mathcal{G}$ is already totally orderd under set inclusion. This contradicts the maximality of $\mathcal{G}$, so there cannot exist an $x \notin W$ such that $W \cup \{x\}$ is orthonormal. Therefore, W is a maximally orthonormal set that contains V.

Q.E.D.

We have shown above that any orthonormal set in an inner product space can be extended to an orthornomal basis for that space. Since any non-trivial inner product space $(H, \langle \cdot, \cdot \rangle)$ has a non-zero element $x \in H$ and $\{\frac{x}{\|x\|}\}$ forms an orthonormal subset of H, the theorem implies that any non-trivial inner product space has a maximally orthonormal subset.

6.5 Lebesgue Decomposition and Radon-Nikodym Theorem

In this section we present the von Neumann proof of the Radon-Nikodym theorem, in which the Lebesgue Decomposition theorem and the Radon-Nikodym theorem are proved in one stroke using Hilbert space methods.

6.5.1 Absolute Continuity and Mutual Singularity

Let $(E, \mathcal{E})$ be a measurable space, and μ, v two measures on $(E, \mathcal{E})$. Recall that v is absolutely continuous with respect to μ, or $v << \mu$, if, for any $A \in \mathcal{E}$, $\mu(A) = 0$ implies $v(A) = 0$. If v is a finite measure, then the following result shows us why the preceding relation is called absolute "continuity":

Lemma 6.13 Let $(E, \mathcal{E})$ be a measurable space, and μ, v two measures on $(E, \mathcal{E})$. If v is finite, then $v << \mu$ if and only if, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $v(A) < \varepsilon$ for any $A \in \mathcal{E}$ such that $\mu(A) < \delta$.

Proof) Suppose first that the $\varepsilon - \delta$ condition above holds, that is, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $v(A) < \varepsilon$ for any $A \in \mathcal{E}$ such that $\mu(A) < \delta$. If $\mu(A) = 0$ for some $A \in \mathcal{E}$, then because $\mu(A) < \delta$, this implies that $v(A) < \varepsilon$ as well. This holds for any $\varepsilon > 0$, so $v(A) = 0$. This in turn holds for any $A \in \mathcal{E}$, so $v << \mu$.

To show that absolute continuity implies the $\varepsilon - \delta$ condition, we proceed by contraposition. Suppose that there exists a $\varepsilon > 0$ such that, for any $\delta > 0$, there exists a $A \in \mathcal{E}$ such that $\mu(A) < \delta$ but $v(A) \geq \varepsilon$. In this case, we can choose a sequence $\{A_n\}_{n \in N_+} \subset \mathcal{E}$ such that $v(A_n) \geq \varepsilon$ and $\mu(A_n) < 2^{-n}$ for any $n \in N_+$. Defining the sequence $\{B_n\}_{n \in N_+}$ as

$$B_n = \bigcup_{i=n}^{\infty} A_i$$

for any $n \in N_+$, $\{B_n\}_{n \in N_+}$ is a decreasing sequence of $\mathcal{E}$-measurable sets. Furthermore, for any $n \in N_+$,

$$\mu(B_n) \leq \sum_{i=n}^{\infty} \mu(A_i) \leq \sum_{i=n}^{\infty} 2^{-i} = 2^{-n+1} \cdot \left(\sum_{i=1}^{\infty} 2^{-i} \right) = 2^{-n+1},$$

which implies that

$$\lim_{n \to \infty} \mu(B_n) = 0,$$

and because $\mu(B_1) \leq 1 < +\infty$, by the sequential continuity of measures,

$$\mu(B) = \lim_{n \to \infty} \mu(B_n) = 0.$$

On the other hand, for any $n \in N_+$,

$$v(B_n) \geq v(A_n) \geq \varepsilon,$$

and because $v(B_1) < +\infty$ by the finiteness assumption, by sequential continuity we have

$$v(B) = \lim_{n \to \infty} v(B_n) \geq \varepsilon.$$

Therefore, $B \in \mathcal{E}$ is a measurable set such that $\mu(B) = 0$ but $v(B) > 0$, which tells us that v is not absolutely continuous with respect to μ.
By contraposition, we can conclude that $v << \mu$ implies the $\varepsilon - \delta$ condition.
Q.E.D.

In section 3.2.4, we defined the indefinite integral of a non-negative measurable function with respect to some measure; specifically, letting μ be a measure on $(E, \mathcal{E})$ and $f \in \mathcal{E}_+$, the indefinite integral v of f with respect to μ was a measure on $(E, \mathcal{E})$ defined as

$$v(A) = \int_A f d\mu$$

for any $A \in \mathcal{E}$. We also saw that, for any $g \in \mathcal{E}_+$,

$$\int_E g dv = \int_E g f d\mu.$$

There, we easily showed that $v << \mu$. The Radon-Nikodym theorem tells us the exact opposite, namely that, if $v << \mu$ and μ, v are σ-finite measures, then v must be the indefinite integral of some non-negative measurable function h with respect to μ, so that the relation

$$v(A) = \int_A h d\mu$$

holds for any $A \in \mathcal{E}$. This function h is called the Radon-Nikodym derivative of v with respect to μ.

A concept closely related to absolute continuity is that of singularity.
Let μ be a measure on $(E,\mathcal{E})$. If there exists a $E_0 \in \mathcal{E}$ such that

$$\mu(E_0 \cap A) = \mu(A)$$

for any $A \in \mathcal{E}$, then we say that μ is concentrated on E_0. Heuristically, this means that μ does not assign a non-zero value to any measurable set outside E_0. Note that, if μ is concentrated on $E_0, E_1 \in \mathcal{E}$, then it is also concentrated on $E_0 \cap E_1$: this can be seen by noting that, for any $A \in \mathcal{E}$,

$$\mu((E_0 \cap E_1) \cap A) = \mu(E_0 \cap (E_1 \cap A)) = \mu(E_1 \cap A) = \mu(A),$$

where we used the fact that μ is concentrated on E_0 and E_1 to justify the second and third equalities.
Now let μ, v be two measures on $(E,\mathcal{E})$. If there exist disjoint measurable sets $E_0, E_1 \in \mathcal{E}$ such that μ is concentrated on E_0 and v on E_1, then we say that the measures μ, v are mutually singular, and denote the relation by $\mu \perp v$.
The following are some properties of mutual singularity and absolute continuity:

Lemma 6.14 Let $(E,\mathcal{E})$ be a measurable space, and μ, v_1, v_2 measures on $(E,\mathcal{E})$. Then, the following hold true:

i) If $v_1 \perp \mu$ and $v_2 \perp \mu$, then $v_1 + v_2 \perp \mu$.

ii) If $v_1 << \mu$ and $v_2 << \mu$, then $v_1 + v_2 << \mu$.

iii) If $v_1 << \mu$ and $v_2 \perp \mu$, then $v_1 \perp v_2$.

iv) If $v_1 << \mu$ and $v_1 \perp \mu$, then $v_1 = 0$ on $\mathcal{E}$.

Proof) i) Suppose that $v_1 \perp \mu$ and $v_2 \perp \mu$. Then, there exist measurable sets E_1, E_2, E_{-1}, E_{-2} such that $E_1 \cap E_{-1} = E_2 \cap E_{-2} = \emptyset$ and

$$v_i(E_i \cap A) = v_i(A), \quad \mu(E_{-i} \cap A) = \mu(A)$$

for $i = 1,2$. Denoting $m = v_1 + v_2$, it then follows that, for any $A \in \mathcal{E}$,

$$m((E_1 \cup E_2) \cap A) = v_1((E_1 \cup E_2) \cap A) + v_2((E_1 \cup E_2) \cap A).$$

Note that

$$\begin{aligned} v_1((E_1 \cup E_2) \cap A) &= v_1((E_1 \cap A) \cup (E_2 \cap A)) \\ &= v_1(E_1 \cap A) + v_1((E_2 \cap A) \setminus (E_1 \cap A)) \\ &= v_1(E_1 \cap A) + v_1((E_2 \setminus E_1) \cap A); \end{aligned}$$

because v_1 is concentrated on E_1, $v_1(E_1^c) = v_1(E_1 \cap E_1^c) = 0$, so that

$$v_1((E_2 \setminus E_1) \cap A) = v_1(E_2 \cap E_1^c \cap A) \leq v_1(E_1^c) = 0$$

and therefore

$$v_1((E_1 \cup E_2) \cap A) = v_1(E_1 \cap A) = v_1(A).$$

Likewise, $v_2((E_1 \cup E_2) \cap A) = v_2(A)$, so we have

$$m((E_1 \cup E_2) \cap A) = v_1(A) + v_2(A) = m(A).$$

This holds for any $A \in \mathcal{E}$, so it follows by definition that m is concentrated on $E_1 \cup E_2$. $E_1 \cup E_2$ and $E_{-1} \cap E_{-2}$ are disjoint, and μ is concentrated on $E_{-1} \cap E_{-2}$, so it follows that $v_1 + v_2 = m \perp \mu$.

ii) Suppose that $v_1 << \mu$ and $v_2 << \mu$. Then, for any $A \in \mathcal{E}$ such that $\mu(A) = 0$, $v_1(A) = v_2(A) = 0$, which implies that $(v_1 + v_2)(A) = 0$ as well. By definition, $v_1 + v_2 << \mu$.

iii) Suppose that $v_1 << \mu$ and $v_2 \perp \mu$. Then, there exist disjoint sets E_0 and E_2 such that v_2 is concentrated on E_2 and μ is concentrated on E_0. By implication, $\mu(E_0^c) = \mu(E_0 \cap E_0^c) = 0$, so by absolute continuity, $v_1(E_0^c) = 0$. This indicates that, for any $A \in \mathcal{E}$,

$$v_1(A) = v_1(E_0 \cap A) + v_1(E_0^c \cap A) = v_1(E_0 \cap A),$$

or that v_1 is concentrated on E_0. Since E_0 and E_2 are disjoint, this implies $v_1 \perp v_2$.

iv) Suppose that $v_1 << \mu$ and $v_1 \perp \mu$. By the preceding result, this implies that $v_1 \perp v_1$, or that there exist disjoint sets E_1, E_2 such that

$$v_1(A) = v_1(E_1 \cap A) = v_1(E_2 \cap A)$$

for any $A \in \mathcal{E}$. By implication,

$$v_1(E_1) = v_1(E_2 \cap E_1) = v_1(\emptyset) = 0,$$

and as such, $v_1(A) = v_1(E_1 \cap A) = 0$ for any $A \in \mathcal{E}$.

Q.E.D.

6.5.2 The Main Theorem

Now we are ready to prove the Lebesgue decomposition and Radon-Nikodym theorems. First, we state some preliminary lemmas:

Lemma 6.15 Let $(E,\mathcal{E})$ be a measurable space, and μ a σ-finite measure on $(E,\mathcal{E})$. Then, there exists a μ-integrable function w taking values in $(0,1)$.

Furthermore, defining $v : \mathcal{E} \to [0,+\infty]$ as

$$v(A) = \int_A w d\mu$$

for any $A \in \mathcal{E}$, or the indefinite integral of w with respect to μ, v is a finite measure on $(E,\mathcal{E})$ such that $\mu << v$.

Proof) Because μ is σ-finite, there exists a measurable partition $\{E_n\}_{n\in N_+} \subset \mathcal{E}$ of E such that $\mu(E_n) < +\infty$ for any $n \in N_+$. Now define the function $w : E \to (0,1)$ as

$$w = \sum_{n=1}^{\infty} \frac{2^{-n}}{1+\mu(E_n)} \cdot I_{E_n};$$

w is well-defined because each summand is non-negative, and it is $\mathcal{E}$-measurable because each I_{E_n} is measurable. Finally, by the MCT for series and the linearity of integration,

$$\int_E w d\mu = \sum_{n=1}^{\infty} \frac{2^{-n}}{1+\mu(E_n)} \cdot \mu(E_n) \leq \sum_{n=1}^{\infty} 2^{-n} = 1 < +\infty,$$

so w is μ-integrable.

Now let v be defined as above. We already showed previously that v is a well-defined measure on $(E,\mathcal{E})$, and it is finite because

$$v(E) = \int_E w d\mu \leq 1 < +\infty.$$

To show that μ is absolutely continuous with respect to v, let $A \in \mathcal{E}$ be a set such that $v(A) = 0$. Then,

$$\int_A w d\mu = \int_E (w \cdot I_A) d\mu = 0,$$

and by the vanishing property of non-negative functions, $w \cdot I_A = 0$ a.e. $[\mu]$, that is,

$$\mu(\{w \cdot I_A > 0\}) = 0.$$

Since w is positive everywhere on E, we can see that $\{w \cdot I_A > 0\} = A$ and as such that $\mu(A) = 0$. By definition, $\mu << v$, and since $v << \mu$ by construction, this actually tells us that v and μ are equivalent measures.

Q.E.D.

The next result shows that, if the average of a complex function over any measurable set is within the interval $[0,1]$, the function itself takes values in $[0,1]$ almost everywhere.

Lemma 6.16 Let $(E,\mathcal{E})$ be a measurable space, and μ a finite measure on $(E,\mathcal{E})$. If, for some complex $\mathcal{E}$-measurable f, we have

$$\frac{1}{\mu(A)}\int_A f d\mu \in [0,1],$$

for any $A\in\mathcal{E}$ with $\mu(A)>0$, then the function f takes values in $[0,1]$ a.e. $[\mu]$.

Proof) We first show that $Im(f)=0$ a.e. $[\mu]$. For any $A\in\mathcal{E}$ such that $\mu(A)>0$,

$$\frac{1}{\mu(A)}\int_A f d\mu = \frac{1}{\mu(A)}\int_A Re(f)d\mu + i\cdot\frac{1}{\mu(A)}\int_A Im(f)d\mu \in [0,1],$$

so

$$\int_A Im(f)d\mu = 0.$$

Suppose that

$$\begin{aligned}\mu(\{Im(f)\neq 0\}) &= \mu\left(\{Im(f)^+>0\}\cup\{Im(f)^->0\}\right)\\ &= \mu\left(\{Im(f)^+>0\}\right)+\mu(\{Im(f)^->0\})>0.\end{aligned}$$

Suppose without loss of generality that $\mu(\{Im(f)^+>0\})>0$. Then, since

$$\{Im(f)^+>0\} = \bigcup_n \underbrace{\{Im(f)^+>1/n\}}_{A_n},$$

by countable subadditivity

$$0<\mu\left(\{Im(f)^+>0\}\right)\leq\sum_{n=1}^{\infty}\mu(A_n).$$

This implies that $\mu(A_n)>0$ for some $n\in N_+$; by design, $\frac{1}{n}\cdot I_{A_n}\leq Im(f)^+\cdot I_{A_n}$, so by the monotonicity of integration,

$$0<\frac{1}{n}\mu(A_n)\leq\int_{A_n} Im(f)^+ d\mu.$$

Since $Im(f)^{-} = 0$ on A_n, it follows that $\int_{A_n} Im(f)^{-} d\mu = 0$, and therefore

$$\int_{A_n} Im(f) d\mu = \int_{A_n} Im(f)^{+} d\mu > 0.$$

This contradicts the fact that $\int_A Im(f) d\mu = 0$ for any $A \in \mathcal{E}$ with $\mu(A) > 0$, so $\mu(\{Im(f) \neq 0\}) = 0$, or in other words, $Im(f) = 0$ a.e. $[\mu]$.

Denoting $Re(f) = g$, we can see that

$$\frac{1}{\mu(A)} \int_A g d\mu \in [0,1]$$

for any $A \in \mathcal{E}$ such that $\mu(A) > 0$. Suppose that $\mu(\{g < 0\}) > 0$. Since

$$\{g < 0\} = \bigcup_n \underbrace{\{g < -1/n\}}_{B_n},$$

and

$$\mu(\{g < 0\}) \leq \sum_{n=1}^{\infty} \mu(B_n),$$

that $\mu(\{g < 0\}) > 0$ indicates that there exists an $n \in N_{+}$ such that $\mu(B_n) > 0$. It then follows that $g \cdot I_{B_n} \leq -\frac{1}{n} \cdot I_{B_n}$, so that

$$\int_{B_n} g d\mu \leq -\frac{1}{n} \mu(B_n) < 0,$$

or

$$\frac{1}{\mu(B_n)} \int_{B_n} g d\mu \leq -\frac{1}{n} < 0$$

by the monotonicity of integration, which contradicts the fact that $\frac{1}{\mu(B_n)} \int_{B_n} g d\mu \geq 0$. Therefore, $\mu(\{g < 0\}) = 0$, that is, $g \geq 0$ a.e. $[\mu]$.

Likewise, because

$$\frac{1}{\mu(A)} \int_A (1-g) d\mu \geq 0$$

for any $A \in \mathcal{E}$ such that $\mu(A) > 0$, $1 - g \geq 0$, or equivalently, $g \leq 1$, a.e. $[\mu]$.
Putting these results together, we can see that $g \in [0,1]$ a.e. $[\mu]$, and because $f = g$ a.e. $[\mu]$, we can conclude that $f \in [0,1]$ a.e. $[\mu]$.

Q.E.D.

The following is the statement and proof of the Lebesgue decomposition and Radon-Nikodym theorems, due to von Neumann:

Theorem 6.17 (The Lebesgue-Radon-Nikodym Theorem)
Let $(E, \mathcal{E})$ be a measurable space, and μ, v σ-finite measures on $(E, \mathcal{E})$. Then,

i) **The Lebesgue Decomposition Theorem**
There exist a unique pair of measures v_a and v_s on $(E, \mathcal{E})$ such that

$$v = v_a + v_s, \quad v_a << \mu, \quad v_s \perp \mu.$$

ii) **The Radon-Nikodym Theorem**
There exists a non-negative $\mathcal{E}$-measurable h such that

$$v_a(A) = \int_A h d\mu$$

for any $A \in \mathcal{E}$. If v_a is finite, then h is unique a.e. $[\mu]$.

Proof) We first prove existence, and then move onto uniqueness. As with most complicated proofs, we proceed in steps:

Step 1: Defining the Auxiliary Measure φ

Suppose initially that v is a finite measure. In this step, we define the auxiliary measure φ as the sum of the finite measure v and a finite measure with respect to which μ is absolutely continuous, constructed using lemma 6.15.

Let the $\mathcal{E}$-measurable positive function w be chosen as in lemma 6.16, so that $\int_E w d\mu = 1$ and w takes values in $(0,1)$. Define the measure φ on $(E, \mathcal{E})$ as

$$\varphi(A) = v(A) + \int_A w d\mu$$

for any $A \in \mathcal{E}$, which is a measure because the mappings $A \mapsto v(A)$ and $A \mapsto \int_A w d\mu$ are measures and the sum of measures is also a measure. Furthermore, because

$$\varphi(E) = v(E) + \int_E w d\mu = v(E) + 1 < +\infty$$

due to the finiteness of v, φ is also a finite measure.

For any $\mathcal{E}$-measurable simple function $f = \sum_{i=1}^{n} \alpha_i \cdot I_{A_i}$,

$$\begin{aligned}\int_E f d\varphi &= \sum_{i=1}^{n} \alpha_i \cdot \varphi(A_i) \\ &= \sum_{i=1}^{n} \alpha_i \cdot \left(v(A_i) + \int_E (w \cdot I_{A_i})\, d\mu \right) \\ &= \int_E f dv + \int_E f w d\mu\end{aligned}$$

by the linearity of integration.

Now let f be a general non-negative $\mathcal{E}$-measurable function and $\{s_n\}_{n \in N_+}$ a sequence of $\mathcal{E}$-measurable non-negative functions increasing to f. Then, by repeated applications of the MCT,

$$\begin{aligned}\int_E f d\varphi &= \lim_{n\to\infty} \int_E s_n d\varphi \\ &= \lim_{n\to\infty} \left(\int_E s_n dv + \int_E (s_n \cdot w) d\mu \right) \\ &= \int_E f dv + \int_E (f \cdot w) d\mu.\end{aligned}$$

By implication, because $\int_E (f \cdot w) d\mu \in [0, +\infty]$ for any $f \in \mathcal{E}_+$,

$$\int_E f dv \leq \int_E f dv + \int_E (f \cdot w) d\mu = \int_E f d\varphi$$

for any $f \in \mathcal{E}_+$.

Step 2: Applying the Riesz Representation Theorem for Hilbert Spaces

In this step, we apply the representation theorem proved in the previous section to express integrals with respect to φ as integrals with respect to v. In effect, we are deriving the Radon-Nikodym derivative of v with respect to φ.

For any $[f]_\varphi \in L^2(\mathcal{E}, \varphi)$,

$$\begin{aligned}\int_E |f| dv \leq \int_E |f| d\varphi = \int_E (|f| \cdot 1)\, d\varphi & \\ \leq \left(\int_E |f|^2 d\varphi \right)^{\frac{1}{2}} \left(\int_E 1 d\varphi \right)^{\frac{1}{2}} & \qquad \text{(Hölder's inequality)} \\ = \|f\|_2 \cdot \varphi(E)^{\frac{1}{2}} < +\infty &\end{aligned}$$

because $\varphi(E) < +\infty$ (φ is a finite measure). Thus, f is v-integrable, and we can define the function $L : L^2(\mathcal{E}, \varphi) \to \mathbb{C}$ as

$$L([f]_\varphi) = \int_E f dv$$

for any $[f]_\varphi \in L^2(\mathcal{E}, \varphi)$. Due to the linearity of integration, L is a linear functional on $L^2(\mathcal{E}, \varphi)$, while the above inequality shows us that

$$|L([f]_\varphi) - L([g]_\varphi)| = \left| \int_E f dv - \int_E g dv \right| \leq \int_E |f - g| dv \leq \|[f]_\varphi - [g]_\varphi\|_2 \cdot \varphi(E)^{\frac{1}{2}}$$

for any $[f]_\varphi, [g]_\varphi \in L^2(\mathcal{E}, \varphi)$. Therefore, L is a continuous linear functional on the Hilbert space $(L^2(\mathcal{E}, \varphi), \langle \cdot, \cdot \rangle_2)$.

By the Riesz representation theorem for Hilbert spaces, there exists a $[g]_\varphi \in L^2(\mathcal{E}, \varphi)$ such that

$$\int_E f dv = L([f]_\varphi) = \langle [f]_\varphi, [g]_\varphi \rangle_2 = \int_E (f\overline{g}) d\varphi$$

for any $[f]_\varphi \in L^2(\mathcal{E}, \varphi)$.
For any $A \in \mathcal{E}$, because $\int_E I_A d\varphi = \varphi(A) < +\infty$ by the finiteness of φ, $[I_A]_\varphi \in L^2(\mathcal{E}, \varphi)$ and therefore

$$v(A) = \int_E I_A dv = \int_E (\overline{g} \cdot I_A)\, d\varphi = \int_A \overline{g} d\varphi.$$

For any $A \in \mathcal{E}$ such that $\varphi(A) > 0$, because

$$v(A) = \int_E I_A dv \leq \int_E I_A d\varphi = \varphi(A),$$

the above equality implies that

$$\frac{1}{\varphi(A)} \int_A \overline{g} d\varphi = \frac{v(A)}{\varphi(A)} \in [0, 1].$$

Since φ is a finite measure, by lemma 6.16, this allows us to conclude that $\overline{g} \in [0, 1]$, or equivalently, $g = \overline{g} \in [0, 1]$ a.e. $[\varphi]$. Thus, we may assume that $g \in [0, 1]$ on E without loss of generality, because g just needs to be an element of the equivalence class $[g]_\varphi$.

Step 3: Defining v_a and v_s

In this step, using the representation derived above, we construct the measures v_a and v_s.

For any bounded $f \in \mathcal{E}_+$, we can now see that

$$\int_E f dv = \int_E (fg) d\varphi = \int_E (fg) dv + \int_E (fgw) d\mu,$$

and by the linearity of integration for integrable complex-valued functions, we have

$$\int_E (1-g) f dv = \int_E (fgw) d\mu.$$

Define $E_a = \{0 \leq g < 1\}$ and $E_s = \{g = 1\}$, and define the measures v_a, v_s by

$$v_a(A) = v(E_a \cap A) \quad \text{and} \quad v_s(A) = v(E_s \cap A)$$

for any $A \in \mathcal{E}$. Then, v_a is concentrated on E_a and v_s is concentrated on E_s, and because E_a, E_s are disjoint measurable sets such that $E_a \cup E_s = E$, for any $A \in \mathcal{E}$

$$v(A) = v(E_a \cap A) + v(E_s \cap A) = v_a(A) + v_s(A),$$

which shows us that $v = v_a + v_s$.

Step 4: Showing that $v_s \perp \mu$

Putting $f = I_{E_s}$ in the above equation yields

$$\int_{E_s} (1-g) dv = 0 = \int_{E_s} gw d\mu = \int_{E_s} w d\mu$$

which implies $\mu(E_s) = 0$ by lemma 6.15. Therefore,

$$\mu(A) = \mu(E_s \cap A) + \mu(E_a \cap A) = \mu(E_a \cap A)$$

for any $A \in \mathcal{E}$, which tells us that μ is concentrated on E_a and thus $v_s \perp \mu$.

Step 5: Showing that $v_a << \mu$

g is a bounded non-negative function, so for any $n \in N_+$ and $A \in \mathcal{E}$, $(1+g+\cdots+g^n)\cdot I_{E_a\cap A}$ is also a bounded non-negative $\mathcal{E}$-measurable function, meaning that

$$\begin{aligned}\int_{E_a\cap A}(1-g^{n+1})dv &= \int_{E_a\cap A}(1-g)(1+g+\cdots+g^n)dv \\ &= \int_{E_a\cap A}(1+g+\cdots+g^n)gwd\mu = \int_{E_a\cap A}(g+\cdots+g^{n+1})wd\mu.\end{aligned}$$

For any $x \in E_a \cap A$, $g(x)^{n+1} \searrow 0$ as $n \to \infty$, so $\{(1-g^{n+1})\cdot I_{E_a\cap A}\}_{n\in N_+}$ is a sequence in $\mathcal{E}_+$ increasing to $I_{E_a\cap A}$, while $\{(g+\cdots+g^{n+1})w\cdot I_{E_a\cap A}\}_{n\in N_+}$ is a sequence in $\mathcal{E}_+$ increasing to $\frac{g}{1-g}w\cdot I_{E_a\cap A}$. Therefore, defining

$$h = \frac{g}{1-g}w\cdot I_{E_a} \in \mathcal{E}_+,$$

which is a real-valued non-negative measurable function, by repeated applications of the MCT, we have

$$\begin{aligned}v_a(A) = v(E_a\cap A) &= \lim_{n\to\infty}\int_{E_a\cap A}(1-g^{n+1})dv \\ &= \lim_{n\to\infty}\int_{E_a\cap A}(g+\cdots+g^{n+1})wd\mu = \int_A hd\mu.\end{aligned}$$

This shows us that h is the Radon-Nikodym derivative of v_a with respect to μ, and as such that $v_a << \mu$.

Step 6: Extending the result to σ-finite measures

Now let v be a σ-finite measure. Then, there exists a partition $\{E_n\}_{n\in N_+}$ of E such that $v(E_n) < +\infty$ for any $n \in N_+$. As in lemma 2.14, for any $n \in N_+$, define the finite measure v_n on $(E,\mathcal{E})$ as $v_n(A) = v(E_n\cap A)$ for any $A \in \mathcal{E}$. By design, each v_n is concentrated on A, and we showed in that lemma that $v = \sum_{n=1}^{\infty} v_n$.

By the preceding result, for any $n \in N_+$, because v_n is a finite measure, there exist measures $v_{a,n}$ and $v_{s,n}$ concentrated on disjoint subsets $E_{a,n}$ and $E_{s,n}$ of E_n whose union is E_n such that

- $v_n = v_{a,n} + v_{s,n}$,
- $v_{s,n} \perp \mu$ (in particular, $\mu(E_{s,n}) = 0$),
- $v_{a,n} << \mu$, and
- There exists a real-valued $h_n \in \mathcal{E}_+$ such that, for any $A \in \mathcal{E}$,

$$v_{a,n}(A) = \int_E h_n d\mu.$$

Then, defining

$$v_a = \sum_{n=1}^{\infty} v_{a,n} \quad \text{and} \quad v_s = \sum_{n=1}^{\infty} v_{s,n},$$

v_a, v_s are measures on $(E, \mathcal{E})$. We show that v_a and v_s satisfy the following conditions:

- $v = v_a + v_s$
  For any $A \in \mathcal{E}$,

$$v(A) = \sum_{n=1}^{\infty} v_n(A) = \sum_{n=1}^{\infty} v_{a,n}(A) + \sum_{n=1}^{\infty} v_{s,n}(A) = v_a(A) + v_s(A).$$

- $v_a \perp v_s$
  Defining $E_a = \bigcup_n E_{a,n}$ and $E_s = \bigcup_n E_{s,n}$, because each E_n is partitioned into $E_{a,n}$ and $E_{s,n}$, and $\{E_n\}_{n \in N_+}$ is itself a partition of E, $\{E_a, E_s\}$ is a partition of E. Note that, for any $A \in \mathcal{E}$,

$$v_a(A) = \sum_{n=1}^{\infty} v_{a,n}(A) = \sum_{n=1}^{\infty} v_{a,n}(E_{a,n} \cap A) = \sum_{n=1}^{\infty} v_{a,n}(E_a \cap A) = v_a(E_a \cap A),$$

  where we used the fact that $E_{a,n} \subset E_a$ and

$$v_{a,n}(E_a \cap A) = v_{a,n}(E_{a,n} \cap A) + v_{a,n}(E_{a,n}^c \cap E_a \cap A) = v_{a,n}(E_{a,n} \cap A)$$

  for any $n \in N_+$ to justify the third equality. This shows us that v_a is concentrated on E_a, and likewise, v_s is concentrated on E_s.

- $v_s \perp \mu$
  Since

$$\mu(E_s) = \sum_{n=1}^{\infty} \mu(E_{s,n}) = 0$$

  by countable additivity, we can see that $v_s \perp \mu$.

- $v_a << \mu$
  Defining $h = \sum_{n=1}^{\infty} h_n$, h is a non-negative $\mathcal{E}$-measurable function such that

$$v_a(A) = \sum_{n=1}^{\infty} v_{a,n}(A) = \sum_{n=1}^{\infty} \int_E (h_n \cdot I_A) d\mu = \int_A h d\mu$$

  by the MCT for series. It follow sthat $v_a << \mu$.

Step 7: Uniqueness of the Radon-Nikodym Derivative for Finite v_a

Now we show that h is unique a.e. $[\mu]$ if v_a is finite. Let $h' \in \mathcal{E}_+$ satisfy

$$\int_A h d\mu = v_a(A) = \int_A h' d\mu$$

for any $A \in \mathcal{E}$. Since v_a is a finite measure, $v_a(E) = \int_E h d\mu = \int_E h' d\mu < +\infty$, meaning that h, h' are μ-integrable and, by the finiteness property of non-negative functions, that there exist real-valued non-negative measurable functions $\bar{h}, \bar{h}'$ that are equivalent to h, h' a.e. $[\mu]$.

Since $\bar{h}, \bar{h}'$ are both real-valued μ integrable functions, the function $k = \bar{h} - \bar{h}'$ is also real-valued and μ integrable. By the linearity of integration, we can see that

$$\int_A k d\mu = \int_A \bar{h} d\mu - \int_A \bar{h}' d\mu = \int_A h d\mu - \int_A h' d\mu = 0$$

for any $A \in \mathcal{E}$. By theorem 3.10, it now follows that $k = 0$, or equivalently, $\bar{h} = \bar{h}'$ a.e. $[\mu]$. Together with the above result, we can conclude that $h = h'$ a.e. $[\mu]$, and as such that h is unique up to a.e. equivalence.

Step 8: Uniqueness of the Lebesgue Decomposition

Suppose there exist measures m_a, m_s such that $m_a << \mu$, $m_s \perp \mu$ and $v = m_a + m_s$. Then,

$$v = v_a + v_s = m_a + m_s.$$

Letting m_a and m_s be concentrated on F_a, F_s, where $F_a \cap F_s = \emptyset$ and $F_a \cup F_s = E$, because $m_s \perp \mu$,

$$\mu(F_s) = \mu(E_s) = 0.$$

Moreover, for any $A \in \mathcal{E}$ such that $\mu(A) = 0$, because $v_a << \mu$ and $m_a << \mu$,

$$\begin{aligned} v(A) &= v_a(A) + v_s(A) = v_s(A) \\ &= m_a(A) + m_s(A) = m_s(A). \end{aligned}$$

Now choose any $A \in \mathcal{E}$. Since $\mu(E_s \cap A) = \mu((F_s \cap E_s) \cap A) = 0$, we have

$$\begin{aligned} v_s(A) &= v_s(E_s \cap A) && (v_s \text{ is concentrated on } E_s) \\ &= m_s(E_s \cap A) && (v_s = m_s \text{ for } \mu\text{-negligible sets}) \\ &= m_s((F_s \cap E_s) \cap A) && (m_s \text{ is concentrated on } F_s) \\ &= v_s((F_s \cap E_s) \cap A). && (v_s = m_s \text{ for } \mu\text{-negligible sets}) \end{aligned}$$

Likewise,

$$m_s(A) = m_s((F_s \cap E_s) \cap A),$$

so it follows that

$$v_s(A) = v_s((F_s \cap E_s) \cap A) = m_s((F_s \cap E_s) \cap A) = m_s(A).$$

This holds for any $A \in \mathcal{E}$, so $v_s = m_s$ on $\mathcal{E}$.

To show that $v_a = m_a$, we proceed similarly. For any $A \in \mathcal{E}$,

$$v_s(E_a \cap A) = m_s(E_a \cap A) = 0,$$

and likewise, $m_s(E_a \cap A) = v_s(F_a \cap A) = 0$.
For any $A \in \mathcal{E}$, we now have

$$\begin{aligned} v_a(A) = v_a(E_a \cap A) = v_a(E_a \cap A) + v_s(E_a \cap A) = v(E_a \cap A) \quad &= m_a(E_a \cap A) + m_s(E_a \cap A) \\ &= m_a(E_a \cap A) = m_a((F_a \cap E_a) \cap A), \end{aligned}$$

and by symmetry,

$$m_a(A) = v_a((F_a \cap E_a) \cap A).$$

Finally, the fact that $v_s((F_a \cap E_a) \cap A) = m_s((F_a \cap E_a) \cap A) = 0$ implies that

$$\begin{aligned} m_a(A) = v_a((F_a \cap E_a) \cap A) &= v((F_a \cap E_a) \cap A) \\ &= m_a((F_a \cap E_a) \cap A) = v_a(A). \end{aligned}$$

This holds for any $A \in \mathcal{E}$, so $v_a = m_a$ on $\mathcal{E}$ and the pair (v_a, v_s) is the unique pair of measures on $(E, \mathcal{E})$ such that $v = v_a + v_s$, $v_a << \mu$ and $v_s \perp \mu$.

Q.E.D.

The familiar form of the Radon-Nikodym theorem, which states that, if $v << \mu$, then there exists a Radon-Nikodym derivative of v with respect to μ, now follows as a special case of the above theorem.

Corollary to the Lebesgue-Radon-Nikodym Theorem
Let $(E, \mathcal{E})$ be a measurable space, and μ, v σ-finite measures on $(E, \mathcal{E})$. If $v << \mu$, then there exists a $h \in \mathcal{E}_+$ such that

$$v(A) = \int_A h d\mu$$

for any $A \in \mathcal{E}$. If, in addition, v is finite, then h is unique a.e. $[\mu]$.

Proof) By the Lebesgue-Radon-Nikodym theorem, there exists a unique pair of measures v_a and v_s on $(E, \mathcal{E})$ such that $v = v_a + v_s$, $v_a << \mu$, and $v_s \perp \mu$. Furthermore, the theorem tells us that there exists an $h \in \mathcal{E}_+$ such that

$$v_a(A) = \int_A h d\mu$$

for any $A \in \mathcal{E}$, and that h is unique up to a.e. equivalence if v_a is finite. The corollary follows if we can just show that $v = v_a$ on $\mathcal{E}$.

Because $v = v_a + v_s << \mu$ by assumption, for any $A \in \mathcal{E}$ such that $\mu(A) = 0$, $v(A) = 0$, and since $v_a << \mu$ as well, we have $v_a(A) = 0$. Therefore,

$$0 = v(A) = v_a(A) + v_s(A) = v_s(A),$$

which tells us that $v_s << \mu$. By lemma 6.14, because $v_2 << \mu$ and $v_s \perp \mu$, $v_2 = 0$ on $\mathcal{E}$, which implies that

$$v(A) = v_a(A) + v_s(A) = v_a(A)$$

for any $A \in \mathcal{E}$.
Q.E.D.

Given σ-finite measures μ, v on $(E, \mathcal{E})$, the set of all Radon-Nikodym derivatives of v with respect to μ is denoted by $\frac{\partial v}{\partial \mu}$. In general, $\frac{\partial v}{\partial \mu}$ is a set of multiple $\mathcal{E}$-measurable non-negative functions h such that

$$v(A) = \int_A h d\mu$$

for any $A \in \mathcal{E}$.
However, if v is finite, since such an h is unique a.e. $[\mu]$, $\frac{\partial v}{\partial \mu}$ becomes an equivalence class of non-negative numerical functions. Furthermore, due to the finiteness of v,

$$v(E) = \int_E h d\mu < +\infty,$$

which, by the finiteness property of non-negative functions, tells us that there exists a real-valued function $f \in \mathcal{E}_+$ that is equal to h a.e. $[\mu]$. Since this f is complex-valued and its integral is finite, we can see that $[f]_\mu \in L^1(\mathcal{E}, \mu)$.
Therefore, in this case we define $\frac{\partial v}{\partial \mu} = [f]_\mu$, which is an element of the Banach space $L^1(\mathcal{E}, \mu)$, and call $\frac{\partial v}{\partial \mu}$ "the" Radon-Nikodym derivative of v with respect to μ.

Chapter 7

Integration on Product Spaces

7.1 Product σ-algebras

Let $(E_1, \mathcal{E}_1), \cdots, (E_n, \mathcal{E}_n)$ be measurable spaces. We define the set of all measurable rectangles as

$$\mathcal{R} = \{A_1 \times \cdots \times A_n \mid A_i \in \mathcal{E}_i \text{ for all } 1 \leq i \leq n\}.$$

$\mathcal{R}$ is then a collection of subsets of the product space $E = E_1 \times \cdots \times E_n$. In particular, $\mathcal{R}$ is a π-system on E, since, for any $A_1 \times \cdots \times A_n, B_1 \times \cdots \times B_n \in \mathcal{R}$, we have

$$(A_1 \times \cdots \times A_n) \cap (B_1 \times \cdots \times B_n) = (A_1 \cap B_1) \times \cdots \times (A_n \cap B_n) \in \mathcal{R},$$

where we used the fact that σ-algebras are closed under finite intersections.
However, $\mathcal{R}$ is not a σ-algebra, since the union of measurable rectangles is not necessarily a measurable rectangle; this can be seen easily by studying open rectangles on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.
We thus define the product σ-algebra $\mathcal{E} = \bigotimes_{i=1}^n \mathcal{E}_i$ as the σ-algebra on E generated by $\mathcal{R}$. This is analogous to how the product topology was defined as the topology generated by the collection of all open rectangles.
The measurable space $(E, \mathcal{E})$ is then called the product of $(E_1, \mathcal{E}_1), \cdots, (E_n, \mathcal{E}_n)$ and detnoed $\bigotimes_{i=1}^n (E_i, \mathcal{E}_i)$.

For any $x \in E$ and $1 \leq i \leq n$, we can write x_{-i} for x without its ith coordinate, and denote $x = (x_i, x_{-i})$. Likewise, we denote $E_{-i} = \prod_{j \neq i} E_j$ and $\mathcal{E}_{-i} = \bigotimes_{j \neq i} \mathcal{E}_j$.
Given any $A \in \mathcal{E}$, the i-section of A at $x_{-i} \in E_{-i}$ is defined as

$$A_{x_{-i}} = \{x \in E_i \mid (x, x_{-i}) \in A\},$$

and likewise, for any set F and function $f : E \to F$, the i-section of f at $x_{-i} \in E_{-i}$ is defined as

$$f_{x_{-i}}(x) = f(x, x_{-i})$$

for any $x \in E_i$.

The following are some properties of product σ-algebras and sections of sets and functions:

Lemma 7.1 Let $(E_1,\mathcal{E}_1),\cdots,(E_n,\mathcal{E}_n)$ be measurable spaces, and $(E,\mathcal{E})$ their product. Then, the following hold true:

i) For any $A \in \mathcal{E}$ and $1 \le i \le n$, the i-section $A_{x_{-i}}$ is $\mathcal{E}_i$-measurable for any $x_{-i} \in E_{-i}$.

ii) Let $(F,\mathcal{F})$ be a measurable space. For any function $f : E \to F$ that is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, the i-section $f_{x_{-i}}$ is measurable relative to $\mathcal{E}_i$ and $\mathcal{F}$ for any $x_{-i} \in E_{-i}$.

Proof) i) Choose any $1 \le i \le n$ and $x_{-i} \in E_{-i}$. Define

$$\mathbb{D} = \{A \subset E \mid A_{x_{-i}} \in \mathcal{E}_i\},$$

which is a collection of subsets of E. It is immediately clear that $\mathbb{D}$ contains all measurable rectangles; for any $A = A_1 \times \cdots \times A_n \in \mathbb{D}$, note that

$$A_{x_{-i}} = \begin{cases} A_i & \text{if } x_{-i} \in \prod_{j \neq i} A_j \\ \emptyset & \text{otherwise} \end{cases} \in \mathcal{E}_i.$$

We can also show that $\mathbb{D}$ is a σ-algebra on E:

- $E \in \mathbb{D}$ because E is a measurable rectangle.
- For any $A \in \mathbb{D}$, because

$$(A^c)_{x_{-i}} = \{x_i \in E_i \mid (x_i,x_{-i}) \in A^c\} = \{x_i \in E_i \mid (x_i,x_{-i}) \notin A\} = \left(A_{x_{-i}}\right)^c,$$

  and $A_{x_{-i}} \in \mathcal{E}_i$, it follows that $(A^c)_{x_{-i}} \in \mathcal{E}_i$ as well, so that $A^c \in \mathbb{D}$.
- For any sequence $\{A_n\}_{n \in N_+}$ in $\mathbb{D}$ with union $A = \bigcup_n A_n$, because

$$A_{x_{-i}} = \{x_i \in E_i \mid (x_i,x_{-i}) \in A\} = \bigcup_n \{x_i \in E_i \mid (x_i,x_{-i}) \in A_n\} = \bigcup_n (A_n)_{x_{-i}}$$

  and $(A_n)_{x_{-i}} \in \mathcal{E}_i$ for each $n \in N_+$, we have $A_{x_{-i}} \in \mathcal{E}_i$ and thus $A \in \mathbb{D}$.

It follows that $\mathbb{D}$ is a σ-algebra containing every measurable rectangle on E. Since $\mathcal{E}$ is defined as the smallest σ-algebra containing the colleciton of measurable rectangles, we have $\mathcal{E} \subset \mathbb{D}$, that is, $A_{x_{-i}} \in \mathcal{E}_i$ for any $A \in \mathcal{E}$.

ii) Let $f : E \to F$ be any function measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and choose any $1 \le i \le n$, $x_{-i} \in E_{-i}$. Define $h : E_i \to E$ as

$$h(x) = (x, x_{-i})$$

for any $x \in E_i$. For any measurable rectangle $A = A_1 \times \cdots \times A_n \in \mathcal{E}$,

$$h^{-1}(A) = \begin{cases} A_i & \text{if } x_{-i} \in \prod_{j \neq i} A_j \\ \emptyset & \text{otherwise} \end{cases}$$

is a $\mathcal{E}_i$-measurable set. The set of all measurable rectangles generates the product σ-algebra $\mathcal{E}$, so by the characterization of measurability, h is measurable relative to $\mathcal{E}_i$ and $\mathcal{E}$.

Since the section $f_{x_{-i}}$ can be expressed as the composition $f \circ h$, where f is measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and h is measurable relative to $\mathcal{E}_i$ and $\mathcal{E}$, $f \circ h$ is measuable relative to $\mathcal{E}_i$ and $\mathcal{F}$.

Q.E.D.

The next result shows us that the product of Borel spaces generated by second countable topological spaces is the Borel space generated by the product of those spaces.

Lemma 7.2 Let $(E_1, \tau_1), \cdots, (E_n, \tau_n)$ be second countable topological spaces with countable bases $\mathbb{B}_1, \cdots, \mathbb{B}_n$ on $E_1, \cdots, E_n$. Then, the Borel σ-algebra generated by the product topology $\tau_1 \times \cdots \times \tau_n$ is exactly the product of the Borel σ-algebras $\mathcal{B}(E_1, \tau_1), \cdots, \mathcal{B}(E_n, \tau_n)$, that is,

$$\mathcal{B}(E_1 \times \cdots \times E_n, \tau_1 \times \cdots \times \tau_n) = \mathcal{B}(E_1, \tau_1) \bigotimes \cdots \bigotimes \mathcal{B}(E_n, \tau_n).$$

Proof) Denote $E = E_1 \times \cdots \times E_n$, $\tau = \tau_1 \times \cdots \times \tau_n$ and $\mathcal{E} = \mathcal{B}(E_1, \tau_1) \otimes \cdots \otimes \mathcal{B}(E_n, \tau_n)$. By definition, $\mathcal{B}(E, \tau)$ is generated by τ and $\mathcal{E}$ by $\mathcal{R}$, the set of all measurable rectangles on E.

Define $\overline{\mathcal{R}}$ as

$$\overline{\mathcal{R}} = \{A_1 \times \cdots \times A_n \mid A_i \in \mathbb{B}_i \text{ for any } 1 \leq i \leq n\},$$

so that $\overline{\mathcal{R}} \subset \mathcal{R}$. Recall that $\overline{\mathcal{R}}$ is a countable base on E generating the product topology τ. By lemma 2.2, this means that $\overline{\mathcal{R}}$ generates $\mathcal{B}(E, \tau)$.

Since $\mathcal{R}$ generates the product σ-algebra $\mathcal{E}$ and $\overline{\mathcal{R}} \subset \mathcal{R}$, it follows that

$$\mathcal{B}(E, \tau) = \sigma\overline{\mathcal{R}} \subset \sigma\mathcal{R} = \mathcal{E}.$$

To show that the reverse inclusion holds, first define the projection function $\pi_i : E \to E_i$ for any $1 \leq i \leq n$ as

$$\pi_i(x_1, \cdots, x_n) = x_i$$

for any $(x_1, \cdots, x_n) \in E$. For any $A \in \tau_i$,

$$\pi_i^{-1}(A) = A \times E_{-i} \in \mathcal{R} \subset \tau,$$

so that π_i is continuous relative to τ and τ_i, and therefore measurable relative to $\mathcal{B}(E, \tau)$ and $\mathcal{B}(E_i, \tau_i)$.
For $A_1 \times \cdots \times A_n \in \mathcal{R}$, we can see that

$$A_1 \times \cdots \times A_n = \bigcap_{i=1}^{n} (A_i \times E_{-i}) = \bigcap_{i=1}^{n} \pi_i^{-1}(A_i);$$

since each $A_i \in \tau_i$, it follows that each $\pi_i^{-1}(A_i) \in \tau$ and therefore

$$A_1 \times \cdots \times A_n \in \tau \subset \mathcal{B}(E, \tau).$$

Therefore, we can see that

$$\mathcal{R} \subset \mathcal{B}(E, \tau),$$

and because $\mathcal{R}$ generates $\mathcal{E}$, we have

$$\mathcal{E} = \sigma\mathcal{R} \subset \mathcal{B}(E, \tau).$$

By implication,

$$\mathcal{E} = \mathcal{B}(E, \tau),$$

and it is clear that, without the second countability condition, we would only have

$$\mathcal{E} \subset \mathcal{B}(E, \tau).$$

Q.E.D.

A corollary to the above theorem is that, for the euclidean n-space $\mathbb{R}^n$, because $(\mathbb{R}, \tau_{\mathbb{R}})$ is a second countable topological space,

$$\mathcal{B}(\mathbb{R}^n, \tau_{\mathbb{R}}^n) = \underbrace{\mathbb{B}(\mathbb{R}) \times \cdots \times \mathbb{B}(\mathbb{R})}_{n}.$$

Furthermore, because the product topology $\tau_{\mathbb{R}}^n$ is the euclidean topology $\tau_{\mathbb{R}^n}$, we have

$$\mathcal{B}(\mathbb{R}^n) = \underbrace{\mathbb{B}(\mathbb{R}) \times \cdots \times \mathbb{B}(\mathbb{R})}_{n} = \mathbb{B}(\mathbb{R})^n.$$

In other words, the Borel σ-algebra on $\mathbb{R}^n$ is the product of n Borel σ-algebras on $\mathbb{R}$.

7.2 Product Measures

So far, we have studied the generation of σ-algebras on product spaces using the σ-algebra on each constituent space. In this section, we study how to create measures on those product spaces. We primarily work with the product of two measurable spaces.

The main workhorse theorem is the following result. First, we define some requisite notation. For measurable spaces $(E,\mathcal{E})$ and $(F,\mathcal{F})$ with product $(E\times F,\mathcal{E}\bigotimes\mathcal{F})$, for any set G and function $f: E\times F\to G$, we denote the 1-section and 2-section of f by

$$f_x(y)=f(x,y) \quad \text{and} \quad f^y(x)=f(x,y)$$

for any $x\in E$, $y\in F$.

A set $A\subset E\times F$ is called an elementary set if there exist disjoint measurable rectangles $A_1,\cdots,A_m\in\mathcal{R}$ such that $A=\bigcup_{i=1}^m A_i$, that is, if it is the union of a finite number of measurable rectangles. We denote the collection of all elementary sets on $E\times F$ by $\mathscr{E}$. It is trivially true that $\mathcal{R}\subset\mathscr{E}$, where $\mathcal{R}$ is the set of all measurable rectangles on $E\times F$.

The following are important properties of $\mathscr{E}$:

Lemma 7.3 Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces, and $\mathscr{E}$ the collection of elementary sets on $E\times F$. Then, the following hold true:

i) $\mathscr{E}$ is an algebra on E

ii) $\mathcal{E}\bigotimes\mathcal{F}$ is the smallest monotone class containing $\mathscr{E}$.

Proof) We first show that $\mathscr{E}$ is an algebra on E.
Clearly, $E\times F\in\mathscr{E}$ because $E\times F$ is a measurable rectangle.

$\mathscr{E}$ is also closed under finite intersections.
To see this, let $A,B\in\mathscr{E}$, where $A_1,\cdots,A_m$ and $B_1,\cdots,B_k$ are disjoint measurable rectangles such that $A=A_1\cup\cdots\cup A_m$ and $B=B_1\cup\cdots\cup B_k$. It then follows that

$$A\cap B=\bigcup_{i=1}^m (A_i\cap B)=\bigcup_{i=1}^m\left(A_i\cap\left(\bigcup_{j=1}^k B_j\right)\right)=\bigcup_{i=1}^m\bigcup_{j=1}^k(A_i\cap B_j).$$

Since $A_i\cap B_j$ are disjoint for distinct (i,j), and the intersection of measurable rectangles is also a measurable rectangle, $A\cap B$ is the finite union of disjoint measurable rectangles and therefore an elementary set.

Finally, we can see that $\mathscr{E}$ is closed under complements.
For any two measurable rectangles $A_1\times B_1, A_2\times B_2\subset E\times F$, note that

$$(A_1\times B_1)\setminus(A_2\times B_2)=((A_1\setminus A_2)\times B_1)\cup((A_1\cap A_2)\times(B_1\setminus B_2)),$$

which is the union of two disjoint measurable rectangles and thus an elementary set. It follows that, for any measurable rectangle $A \in \mathcal{R}$, since $A^c = (E \times F) \setminus A$, $A^c \in \mathscr{E}$.
Now let $A \in \mathscr{E}$, so that there exist disjoint measurable rectangles $A_1, \cdots, A_m$ on $E \times F$ such that $A = A_1 \cup \cdots \cup A_m$. Then,

$$A^c = A_1^c \cap \cdots \cap A_2^c,$$

where $A_1^c, \cdots, A_2^c \in \mathscr{E}$. Since we showed above that the finite intersection of elementary sets is elementary, this means that $A^c \in \mathscr{E}$, and as such that $\mathscr{E}$ is closed under complements.

For any $A, B \in \mathscr{E}$, we can now see that $A \cup B = (A^c \cap B^c)^c \in \mathscr{E}$, since $\mathscr{E}$ is closed under both complements and finite intersections. As such, $\mathscr{E}$ is closed under finite unions as well, which tells us that $\mathscr{E}$ is an algebra on E.

Now we need only apply the monotone class theorem to derive the second result.
The collection $\mathcal{R}$ of measurable rectangles on $E \times F$ generates $\mathcal{E} \bigotimes \mathcal{F}$, and $\mathcal{R}$ is contained in the collection $\mathscr{E}$ of elementary sets on E, so the σ-algebra generated by $\mathscr{E}$ contains the product σ-algebra $\mathcal{E} \bigotimes \mathcal{F}$; $\mathcal{E} \bigotimes \mathcal{F} \subset \sigma\mathscr{E}$.
On the other hand, because $\mathscr{E}$ is the collection of finite disjoint unions of measurable rectangles, measurable rectangles are contained in $\mathcal{E} \bigotimes \mathcal{F}$, and $\mathcal{E} \bigotimes \mathcal{F}$ is closed under finite unions, $\mathscr{E} \subset \mathcal{E} \bigotimes \mathcal{F}$. This implies that the σ-algebra generated by $\mathscr{E}$ is contained in $\mathcal{E} \bigotimes \mathcal{F}$, and together with the preceding result, we can conclude that $\mathcal{E} \bigotimes \mathcal{F}$ is the σ-algebra generated by $\mathscr{E}$. Since $\mathscr{E}$ is an algebra on E, it now follows that by the monotone class theorem that $\mathcal{E} \bigotimes \mathcal{F}$ is the smallest monotone class containing $\mathscr{E}$.
Q.E.D.

Theorem 7.4 Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},v)$ be σ-finite measure spaces. Then, for any $A \in \mathcal{E} \bigotimes \mathcal{F}$, defining

$$\varphi(x) = \int_F (I_A)_x dv \quad \text{and} \quad \psi(y) = \int_E (I_A)^y d\mu$$

for any $x \in E$ and $y \in F$, φ and ψ are $\mathcal{E}$- and $\mathcal{F}$-measurable functions taking values in $[0,+\infty]$, and

$$\int_E \varphi d\mu = \int_F \psi dv.$$

Proof) By σ-finiteness, there exist measurable partitions $\{E_n\}_{n\in N_+}$ and $\{F_n\}_{n\in N_+}$ of E and F such that

$$\mu(E_n), v(F_n) < +\infty$$

for any $n \in N_+$.

From previous results we know that the sections $(I_A)_x$ and $(I_A)^y$ defined as

$$(I_A)_x(y) = I_A(x,y) \quad \text{and} \quad (I_A)^y = I_A(x,y)$$

are both non-negative $\mathcal{F}$- and $\mathcal{E}$-measurable functions for any $x \in E$ and $y \in F$. Then,

$$\varphi(x) = \int_F (I_A)_x dv \quad \text{and} \quad \psi(y) = \int_E (I_A)^y d\mu$$

are both well-defined in $[0,+\infty]$, for any $x \in E$ and $y \in F$.

We now proceed in steps.

Step 1: The Case for Measurable Rectangles and Elementary Sets

Throughout, for any $A \subset E \times F$, let φ_A and ψ_A be the functions on E and F defined as

$$\varphi_A(x) = \int_F (I_A)_x dv$$
$$\psi_A(y) = \int_E (I_A)^y d\mu$$

for any $x \in E$, $y \in F$.
For any $n, m \in N_+$, denote $\Omega_{nm} = E_n \times F_m$, and let $\mathcal{M}$ be the collection of subsets of A of $E \times F$ that satisfy the following conditions:

$$A \in \mathcal{E} \bigotimes \mathcal{F}$$
$$\varphi_{A \cap \Omega_{nm}} \in \mathcal{E}_+$$
$$\psi_{A \cap \Omega_{nm}} \in \mathcal{F}_+$$
$$\int_E \varphi_{A \cap \Omega_{nm}} d\mu = \int_F \psi_{A \cap \Omega_{nm}} dv.$$

It is immediately clear that $\mathcal{M}$ contains every measurable rectangle on $E \times F$; for any measurable rectangle $A \times B \subset E \times F$, because

$$I_{(A \times B) \cap \Omega_{nm}} = I_{(A \cap E_n) \times (B \cap F_m)},$$

$$(I_{(A \times B) \cap \Omega_{nm}})_x(y) = (I_{(A \times B) \cap \Omega_{nm}})^y(x) = I_{A \cap E_n}(x) I_{B \cap F_m}(y)$$

for any $x \in E$ and $y \in F$, we have

$$\varphi_{(A \times B) \cap \Omega_{nm}}(x) = I_{A \cap E_n}(x) \cdot v(B \cap F_m)$$
$$\psi_{(A \times B) \cap \Omega_{nm}}(y) = \int_E (I_{A \times B})^y d\mu = I_{B \cap F_m}(y) \cdot \mu(A \cap E_n)$$

for any $x \in E$ and $y \in F$. Therefore, $\varphi_{(A \times B) \cap \Omega_{nm}} \in \mathcal{E}_+$, $\psi_{(A \times B) \cap \Omega_{nm}} \in \mathcal{F}_+$ and

$$\int_E \varphi_{(A \times B) \cap \Omega_{nm}} d\mu = \mu(A \cap E_n) v(B \cap F_m) = \int_F \psi_{(A \times B) \cap \Omega_{nm}} dv.$$

By definition, $A \times B \in \mathcal{M}$.

We can also show that $\mathcal{M}$ is closed under finite disjoint unions. Letting $A, B \in \mathcal{M}$ be disjoint, note that

$$I_{(A \cup B) \cap \Omega_{nm}} = I_{A \cap \Omega_{nm}} + I_{B \cap \Omega_{nm}}$$

because $A \cap \Omega_{nm}$ and $B \cap \Omega_{nm}$ are disjoint, and as such the sections of $I_{(A \cup B) \cap \Omega_{nm}}$ are the sum of the sections of $I_{A\cap\Omega_{nm}}$ and $I_{B\cap\Omega_{nm}}$. By implication,

$$\begin{aligned}\varphi_{(A\cup B)\cap\Omega_{nm}} &= \int_F (I_{(A\cup B)\cap\Omega_{nm}})_x dv \\ &= \int_F (I_{A\cap\Omega_{nm}})_x dv + \int_F (I_{B\cap\Omega_{nm}})_x dv = \varphi_{A\cap\Omega_{nm}} + \varphi_{B\cap\Omega_{nm}}\end{aligned}$$

by the linearity of integration and likewise,

$$\psi_{(A\cup B)\cap\Omega_{nm}} = \psi_{A\cap\Omega_{nm}} + \psi_{B\cap\Omega_{nm}}.$$

It follows that $\varphi_{(A\cup B)\cap\Omega_{nm}} \in \mathcal{E}_+$, $\psi_{(A\cup B)\cap\Omega_{nm}} \in \mathcal{F}_+$ and

$$\begin{aligned}\int_E \varphi_{(A\cup B)\cap\Omega_{nm}} d\mu &= \int_E \varphi_{A\cap\Omega_{nm}} d\mu + \int_E \varphi_{B\cap\Omega_{nm}} d\mu \\ &= \int_F \psi_{A\cap\Omega_{nm}} dv + \int_F \psi_{B\cap\Omega_{nm}} dv = \int_F \psi_{(A\cup B)\cap\Omega_{nm}} dv.\end{aligned}$$

Therefore, $A \cup B \in \mathcal{M}$.

Since every measurable rectangle is contained in $\mathcal{M}$, and any elementary set is the finite union of disjoint measurable rectangles, the above results tell us that $\mathcal{M}$ contains every elementary set on $E \times F$.

Step 2: The Case for Arbitrary $\mathcal{E}\bigotimes\mathcal{F}$-Measurable Sets

Finally, we will show that $\mathcal{M}$ is a monotone class of sets on $E\times F$.

i) For any increasing sequence $\{A_k\}_{k\in N_+}$ in $\mathcal{M}$ with limit A, note that, for any $k\in N_+$, because $A_k\cap\Omega_{nm}\subset A_{k+1}\cap\Omega_{nm}$ and $\bigcup_k(A_k\cap\Omega_{nm})=A\cap\Omega_{nm}$, we have

$$I_{A_k\cap\Omega_{nm}}\leq I_{A_{k+1}\cap\Omega_{nm}}$$

on $E\times F$ and

$$\lim_{k\to\infty}I_{A_k\cap\Omega_{nm}}=I_{A\cap\Omega_{nm}}.$$

This implies that $\{I_{A_k\cap\Omega_{nm}}\}_{k\in N_+}$ is a sequence of non-negative $\mathcal{E}\bigotimes\mathcal{F}$-measurable functions increasing to $I_{A\cap\Omega_{nm}}$ (where the measurability follows because each A_k is measurable by hypothesis), and as such that $\{(I_{A_k\cap\Omega_{nm}})_x\}_{k\in N_+}$ is a sequence of non-negative $\mathcal{F}$-measurable functions increasing to $(I_{A\cap\Omega_{nm}})_x$ for any $x\in E$. By the monotonicity of integration and the MCT, it follows that

$$\begin{aligned}\varphi_{A\cap\Omega_{nm}}(x)&=\int_F(I_{A\cap\Omega_{nm}})_x dv\\&=\lim_{k\to\infty}\int_F(I_{A_k\cap\Omega_{nm}})_x dv=\lim_{k\to\infty}\varphi_{A_k\cap\Omega_{nm}}(x)\end{aligned}$$

and that

$$\varphi_{A_k\cap\Omega_{nm}}(x)\leq\varphi_{A_{k+1}\cap\Omega_{nm}}(x)$$

for any $x\in E$. Therefore, $\{\varphi_{A_k\cap\Omega_{nm}}\}_{k\in N_+}$ is an increasing sequence of non-negative $\mathcal{E}$-measurable functions (where the measurability follows by hypothesis) with limit $\varphi_{A\cap\Omega_{nm}}$.

As such, $\varphi_{A\cap\Omega_{nm}}\in\mathcal{E}_+$ by the preservation of measurability across limits, and by the MCT,

$$\int_E\varphi_{A\cap\Omega_{nm}}d\mu=\lim_{k\to\infty}\int_E\varphi_{A_k\cap\Omega_{nm}}d\mu.$$

By a symmetric argument, $\{\psi_{A_k\cap\Omega_{nm}}\}_{k\in N_+}$ is also a sequence of non-negative $\mathcal{F}$-measurable functions increasing to $\psi_{A\cap\Omega_{nm}}$ which satisfy

$$\int_F\psi_{A\cap\Omega_{nm}}dv=\lim_{k\to\infty}\int_F\psi_{A_k\cap\Omega_{nm}}dv$$

by the MCT. Since

$$\int_E\varphi_{A_k\cap\Omega_{nm}}=\int_F\psi_{A_k\cap\Omega_{nm}}dv$$

for any $k \in N_+$ by the assumption that $A_k \in \mathcal{M}$, we have

$$\begin{aligned}\int_E \varphi_{A \cap \Omega_{nm}} d\mu &= \lim_{k \to \infty} \int_E \varphi_{A_k \cap \Omega_{nm}} d\mu \\ &= \lim_{k \to \infty} \int_F \psi_{A_k \cap \Omega_{nm}} dv = \int_F \psi_{A \cap \Omega_{nm}} dv.\end{aligned}$$

By definition, $A \in \mathcal{M}$, so $\mathcal{M}$ is closed under increasing limits.

ii) Now let $\{A_k\}_{k \in N_+}$ be a decreasing sequence in $\mathcal{M}$ with limit A. In this case, by the same reasoning as in the above case for increasing sequences of sets in $\mathcal{M}$, $\{(I_{A_k \cap \Omega_{nm}})_x\}_{k \in N_+}$ is a sequence of non-negative $\mathcal{E}$-measurable functions decreasing to $(I_{A \cap \Omega_{nm}})_x$ for any $x \in E$.

Because all of the functions in the above sequence are bounded above by $(I_{\Omega_{nm}})_x = I_{E_n}(x) \cdot I_{F_m}$, and

$$\int_F (I_{\Omega_{nm}})_x dv = v(F_m) I_{E_n}(x) < +\infty$$

because $v(F_m) < +\infty$ by design, by the DCT we have

$$\begin{aligned}\varphi_{A \cap \Omega_{nm}}(x) &= \int_F (I_{A \cap \Omega_{nm}})_x dv \\ &= \lim_{k \to \infty} \int_F (I_{A_k \cap \Omega_{nm}})_x dv = \lim_{k \to \infty} \varphi_{A_k \cap \Omega_{nm}}(x).\end{aligned}$$

By the montonicity of integration, it follows that $\{\varphi_{A_k \cap \Omega_{nm}}\}_{k \in N_+}$ is a decreasing sequence of non-negative $\mathcal{E}$-measurable functions (where the measurability follows because each $A_k \in \mathcal{M}$) with limit $\varphi_{A \cap \Omega_{nm}}$, where each function in $\{\varphi_{A_k \cap \Omega_{nm}}\}_{k \in N_+}$ is bounded above by

$$\varphi_{\Omega_{nm}} = v(F_m) I_{E_n},$$

whose integral with respect to μ is

$$\int_E \varphi_{\Omega_{nm}} d\mu = v(F_m) \mu(E_n) < +\infty.$$

By the DCT again, we have $\varphi_{A \cap \Omega_{nm}} \in \mathcal{E}_+$ and

$$\int_E \varphi_{A \cap \Omega_{nm}} d\mu = \lim_{k \to \infty} \int_E \varphi_{A_k \cap \Omega_{nm}} d\mu.$$

By a symmetric argument for the 1-sections, we can conclude that $\{\psi_{A_k \cap \Omega_{nm}}\}_{k \in N_+}$ is a decreasing sequence of non-negative $\mathcal{F}$-measurable functions with limit $\psi_{A \cap \Omega_{nm}} \in$

$\mathcal{F}_+$, where

$$\int_F \psi_{A\cap\Omega_{nm}} dv = \lim_{k\to\infty} \int_F \psi_{A_k\cap\Omega_{nm}} dv.$$

Because each A_k is contained in $\mathcal{M}$,

$$\int_E \varphi_{A_k\cap\Omega_{nm}} d\mu = \int_F \psi_{A_k\cap\Omega_{nm}} dv,$$

which implies that

$$\begin{aligned}\int_E \varphi_{A\cap\Omega_{nm}} d\mu &= \lim_{k\to\infty} \int_E \varphi_{A_k\cap\Omega_{nm}} d\mu \\ &= \lim_{k\to\infty} \int_F \psi_{A_k\cap\Omega_{nm}} dv = \int_F \psi_{A\cap\Omega_{nm}} dv.\end{aligned}$$

Therefore, by definition $A \in \mathcal{M}$, and $\mathcal{M}$ is closed under decreasing limits.

$\mathcal{M}$ is a monotone class of sets containing $\mathscr{E}$, the set of all elementary sets on $E \times F$. Because $\mathcal{E} \bigotimes \mathcal{F}$ is the smallest monotone class containing $\mathscr{E}$, this implies that $\mathcal{E} \times \mathcal{F} \subset \mathcal{M}$, that is,

$$\begin{aligned}\varphi_{A\cap\Omega_{nm}} &\in \mathcal{E}_+ \\ \psi_{A\cap\Omega_{nm}} &\in \mathcal{F}_+ \\ \int_E \varphi_{A\cap\Omega_{nm}} d\mu &= \int_F \psi_{A\cap\Omega_{nm}} dv.\end{aligned}$$

for any $A \in \mathcal{E} \bigotimes \mathcal{F}$.

Step 3: Extending beyond $E_n \times F_m$

Now choose any $A \in \mathcal{E} \bigotimes \mathcal{F}$. Note that

$$A = \bigcup_n \bigcup_m (A \cap \Omega_{nm}),$$

where $\{A \cap \Omega_{nm} = B_{nm}\}_{n,m \in N_+}$ is a sequence of disjoint subsets of $E \times F$. We showed above that, for any $n, m \in N+$,

$$\begin{aligned} \varphi_{B_{nm}} &\in \mathcal{E}_+ \\ \psi_{B_{nm}} &\in \mathcal{F}_+ \\ \int_E \varphi_{B_{nm}} d\mu &= \int_F \psi_{B_{nm}} dv. \end{aligned}$$

Since

$$I_A = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} I_{B_{nm}}$$

on $E \times F$, we have

$$(I_A)_x = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (I_{B_{nm}})_x$$

for any $x \in E$, and as such, by the MCT for series,

$$\varphi_A(x) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_F (I_{B_{nm}})_x dv = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \varphi_{B_{nm}}(x).$$

By the MCT for series, we can now say that $\varphi_A \in \mathcal{E}_+$ and

$$\int_E \varphi_A d\mu = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_E \varphi_{B_{nm}} d\mu.$$

By a symmetric argument, it holds that $\psi_A \in \mathcal{F}_+$ and

$$\int_F \psi_A dv = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_F \psi_{B_{nm}} dv,$$

so we have

$$\begin{aligned} \int_E \varphi_A d\mu &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_E \varphi_{B_{nm}} d\mu \\ &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \int_F \psi_{B_{nm}} dv = \int_F \psi_A dv. \end{aligned}$$

This holds for any $A \in \mathcal{E} \bigotimes \mathcal{F}$, so the proof is complete.
Q.E.D.

For any $A \in \mathcal{E} \bigotimes \mathcal{F}$, we denote

$$\int_F I_A(x,y)dv(y) := \varphi(x)$$
$$\int_E I_A(x,y)d\mu(x) := \psi(y)$$

for any $x \in E$, $y \in F$ and define

$$\int_E \int_F I_A(x,y)dv(y)d\mu(x) := \int_E \varphi d\mu$$
$$\int_F \int_E I_A(x,y)d\mu(x)dv(y) := \int_F \psi dv.$$

The above theorem tells us that

$$\int_F I_A(\cdot,y)dv(y) \in \mathcal{E}_+, \quad \int_E I_A(x,\cdot)d\mu(x) \in \mathcal{F}_+$$

and

$$\int_E \int_F I_A(x,y)dv(y)d\mu(x) = \int_F \int_E I_A(x,y)d\mu(x)dv(y).$$

for any $A \in \mathcal{E} \bigotimes \mathcal{F}$.

The product measure $\mu \times v$ on $(E \times F, \mathcal{E} \bigotimes \mathcal{F})$ is defined as

$$(\mu \times v)(A) = \int_E \int_F I_A(x,y) dv(y) d\mu(x) = \int_F \int_E I_A(x,y) d\mu(x) dv(y)$$

for any $A \in \mathcal{E} \bigotimes \mathcal{F}$.
Clearly, $(\mu \times v)(\emptyset) = 0$, and for any disjoint sequence of measurable sets $\{A_n\}_{n \in N_+}$ in $\mathcal{E} \bigotimes \mathcal{F}$, we have

$$I_A = \sum_{n=1}^{\infty} I_{A_n}$$

on $E \times F$ and thus, for any $x \in E$,

$$\int_F I_A(x,y) dv(y) = \sum_{n=1}^{\infty} \int_F I_{A_n}(x,y) dv(y)$$

by the MCT for series. Likewise, because $\{\int_F I_{A_n}(\cdot, y) dv(y)\}_{n \in N_+}$ is a sequence of non-negative $\mathcal{E}$-measurable functions, by the MCT for series again we have

$$\begin{aligned}(\mu \times v)(A) &= \int_E \int_F I_A(x,y) dv(y) d\mu(x) \\ &= \int_E \left[\sum_{n=1}^{\infty} \int_F I_{A_n}(x,y) dv(y)\right] d\mu(x) = \sum_{n=1}^{\infty} \int_E \int_F I_{A_n}(x,y) dv(y) d\mu(x) = \sum_{n=1}^{\infty} (\mu \times v)(A_n).\end{aligned}$$

This shows us that $\mu \times v$ is countable additive and therefore a measure on $(E \times F, \mathcal{E} \bigotimes \mathcal{F})$.
For any measurable rectangle $A \times B$, we now have

$$(\mu \times v)(A \times B) = \int_E \left(I_A(x) \cdot \int_F I_B(y) dv(y) \right) d\mu(x) = \mu(A) v(B).$$

Note also that, becuase μ and v are σ-finite, $\mu \times v$ is also σ-finite. To see this, let $\{E_n\}_{n \in N_+}$ and $\{F_m\}_{m \in N_+}$ be partitions of E and F defined in the proof above. It follows that

$$(\mu \times v)(E_n \times F_m) = \mu(E_n) v(F_m) < +\infty$$

for any $n, m \in N_+$, and $\{E_n \times F_m\}_{n,m \in N_+}$ is a measurable partition of $E \times F$, so by definition $\mu \times v$ is σ-finite.

7.3 Fubini's Theorem

In the previous section, given σ-finite measure spaces $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},v)$, we have defined a σ-finite measure space $(E\times F,\mathcal{E}\bigotimes\mathcal{F},\mu\times v)$, where $\mu\times v$ is well-defined as

$$(\mu\times v)(A)=\int_E\int_F I_A(x,y)dv(y)d\mu(x)=\int_F\int_E I_A(x,y)d\mu(x)dv(y)$$

for any $A\in\mathcal{E}\bigotimes\mathcal{F}$.
It is immediately clear that the leftmost term can be interpreted as the integral of the indicator I_A with respect to the product measure $\mu\times v$, and that the two terms on the right show that this integral can be computed via integrals with respect to μ and v, where the order of integration is immaterial. The objective of this section is to extend this result to arbitrary non-negative and complex integrable functions on $E\times F$.
The statement and proof of the theorem are given in the following:

Theorem 7.5 (Fubini's Theorem)
Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},v)$ be σ-finite measure spaces, and f a numerical or complex-valued $\mathcal{E}\bigotimes\mathcal{F}$-measurable function on $E\times F$. Then, the following hold true:

i) If f takes values in $[0,+\infty]$, then the functions $\varphi:E\to[0,+\infty]$, $\psi:F\to[0,+\infty]$ defined as

$$\varphi(x)=\int_F f_x dv \quad\text{and}\quad \psi(y)=\int_E f^y d\mu$$

for any $(x,y)\in E\times F$ are measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and

$$\int_{E\times F} fd(\mu\times v)=\int_E\varphi d\mu=\int_F\psi dv.$$

ii) If f is complex valued and

$$\int_E\varphi^* d\mu<+\infty$$

for the function $\varphi^*:E\to[0,+\infty]$ defined as

$$\varphi^*(x)=\int_F |f|_x dv$$

for any $x\in E$, then f is $\mu\times v$-integrable.

iii) If f is complex valued and $\mu\times v$-integrable, then f_x is v-integrable for μ-a.e. $x\in E$, f^y is μ-integrable for v-a.e. $y\in F$, and for the functions $\varphi:E\to\mathbb{C}$, $\psi:F\to\mathbb{C}$ defined as

$$\varphi(x)=\int_F f_x dv \quad\text{and}\quad \psi(y)=\int_E f^y d\mu$$

for almost every $x \in E$ and $y \in F$, φ and ψ are measurable relative to $\mathcal{E}$ and $\mathcal{F}$, and

$$\int_{E\times F} fd(\mu \times v) = \int_E \varphi d\mu = \int_F \psi dv \in \mathbb{C}.$$

Proof) We again proceed in steps.

Step 1: Indicator Functions

Suppose $f = I_A$ for some $A \in \mathcal{E} \bigotimes \mathcal{F}$. Then, defining

$$\varphi(x) = \int_F f_x dv \quad \text{and} \quad \psi(y) = \int_E f^y d\mu$$

for any $x \in E$, $y \in F$, by the definition of the product measure

$$\begin{aligned}\int_{E\times F} fd(\mu \times v) &= (\mu \times v)(A) \\ &= \int_E \varphi d\mu = \int_F \psi dv,\end{aligned}$$

where the last equality follows from theorem 7.4.

Step 2: Simple Functions

Now let f be a measurable simple function on $E \times F$ with canonical form

$$f = \sum_{i=1}^n \alpha_i \cdot I_{A_i}$$

for $\alpha_1, \cdots, \alpha_n \in [0, +\infty)$ and $A_1, \cdots, A_n \in \mathcal{E} \bigotimes \mathcal{F}$. Define $\varphi_i : E \to [0, +\infty]$ as

$$\varphi_i(x) = \int_F (I_{A_i})_x dv$$

for any $x \in E$ and $1 \leq i \leq n$. We showed above that $\varphi_i \in \mathcal{E}_+$ and that

$$(\mu \times v)(A_i) = \int_F \varphi_i d\mu.$$

For any $x \in E$, we have

$$f_x = \sum_{i=1}^n \alpha_i \cdot (I_{A_i})_x,$$

so that

$$\varphi(x) = \int_F f_x dv = \sum_{i=1}^{n} \alpha_i \cdot \int_F (I_{A_i})_x dv = \sum_{i=1}^{n} \alpha_i \cdot \varphi_i(x),$$

implying that φ, being the linear combination of non-negative measurable functions, is also measurable. Furthermore,

$$\begin{aligned}\int_{E\times F} f d(\mu \times v) = \sum_{i=1}^{n} \alpha_i \cdot (\mu \times v)(A_i) &= \sum_{i=1}^{n} \alpha_i \cdot \left(\int_E \varphi_i d\mu \right) \\ &= \int_E \left(\sum_{i=1}^{n} \alpha_i \cdot \varphi_i \right) d\mu(x) = \int_E \varphi d\mu\end{aligned}$$

by the linearity of integration.

By a symmetric argument, it follows that $\psi \in \mathcal{F}_+$ and

$$\int_{E\times F} f d(\mu \times v) = \int_F \psi dv.$$

Step 3: Non-negative Functions

Finally, let $f \in (\mathcal{E} \bigotimes \mathcal{F})_+$ in general. Then, letting $\{f_n\}_{n\in N_+}$ be an increasing sequence of $\mathcal{E} \bigotimes \mathcal{F}$-measurable simple functions, define $\varphi_n : E \to [0, +\infty]$ as

$$\varphi_n(x) = \int_F (f_n)_x dv$$

for any $x \in E$ and $n \in N_+$. Because each f_n is simple, we saw above that $\varphi_n \in \mathcal{E}_+$ and

$$\int_{E\times F} f_n d(\mu \times v) = \int_E \varphi_n d\mu.$$

For any $x \in E$, $\{(f_n)_x\}_{n\in N_+}$ is an increasing sequence of non-negative $\mathcal{F}$-measurable functions with limit f_x. It then follows from the MCT that

$$\varphi(x) = \int_F f_x dv = \lim_{n\to\infty} \int_F (f_n)_x dv = \lim_{n\to\infty} \varphi_n(x).$$

Becausae φ is the pointwise limit of a sequence of non-negative $\mathcal{E}$-measurable functions, φ is itself non-negative $\mathcal{E}$-measurable. Furthermore, for any $n \in N_+$, by the monotonicity of integration

$$\varphi_n(x) = \int_F (f_n)_x dv \leq \int_F (f_{n+1})_x dv = \varphi_{n+1}(x)$$

for any $x \in E$, meaning that $\{\varphi_n\}_{n\in N_+}$ is a sequence of functions in $\mathcal{E}_+$ increasing to

φ.
By the MCT again,

$$\int_{E\times F} f d(\mu\times v) = \lim_{n\to\infty}\int_{E\times F} f_n d(\mu\times v)$$
$$= \lim_{n\to\infty}\int_E \varphi_n d\mu = \int_E \varphi d\mu.$$

By a symmetric argument, it follows that $\psi\in\mathcal{F}_+$ and

$$\int_{E\times F} f d(\mu\times v) = \int_F \psi dv.$$

This proves the first part of the theorem.

Step 4: Real-valued Functions

Now let f be real-valued and assume that

$$\int_E \varphi^* d\mu < +\infty$$

for $\varphi^* : E\to[0,+\infty]$ defined as

$$\varphi^*(x) = \int_F |f|_x dv$$

for any $x\in E$. Becauase $|f| = f^+ + f^-$, we have $|f|_x = f_x^+ + f_x^-$ for any $x\in E$, which implies that

$$\varphi^*(x) = \int_F |f|_x dv = \int_F f_x^+ dv + \int_F f_x^- dv$$

for any $x\in E$. Define

$$\varphi_\pm(x) = \int_F f_x^\pm dv$$

for any $x\in E$, so that $\varphi^* = \varphi_+ + \varphi_-$. By the preceding result, $\varphi_\pm\in\mathcal{E}_+$, and

$$\int_{E\times F} f^\pm d(\mu\times v) = \int_E \varphi_\pm d\mu.$$

We can now see that

$$\int_E \varphi^* d\mu = \int_E(\varphi_+ + \varphi_-)d\mu = \int_E \varphi_+ d\mu + \int_E \varphi_- d\mu$$
$$= \int_{E\times F} f^+ d(\mu\times v) + \int_{E\times F} f^- d(\mu\times v) = \int_{E\times F}|f| d(\mu\times v).$$

Since $\int_E \varphi d\mu < +\infty$, we have

$$\int_{E\times F} |f| d(\mu \times v) < +\infty,$$

and as such f is $\mu \times v$-integrable.

Conversely, if f is $\mu \times v$-integrable, then

$$\int_{E\times F} f^{\pm} d(\mu \times v) < +\infty.$$

Defining $\varphi_{\pm} \in \mathcal{E}_+$ as above, this means that

$$\int_E \varphi_{\pm} d\mu < +\infty,$$

so by the finiteness property,

$$\mu(\{\varphi_+ = +\infty\}) = \mu(\{\varphi_- = +\infty\}) = 0.$$

This indicates that

$$\mu(\{\varphi_+ = +\infty\} \cup \{\varphi_- = +\infty\}) = 0$$

as well, or that

$$\varphi_+(x) = \int_F (f^+)_x dv < +\infty \quad \text{and} \quad \varphi_-(x) = \int_F (f^-)_x dv < +\infty$$

for μ-a.e. $x \in E$. Because $f_x = (f^+)_x - (f^-)_x$ for any $x \in E$, this means that f_x is v-integrable for μ-a.e. $x \in E$.

It then follows that

$$\varphi(x) = \int_F f_x dv = \int_F (f^+)_x dv - \int_F (f^-)_x dv = \varphi_+(x) - \varphi_-(x)$$

is well-defined for μ-a.e. $x \in E$, and that

$$\begin{aligned}\int_E \varphi d\mu &= \int_E \varphi_+ d\mu - \int_E \varphi_- d\mu \\ &= \int_{E\times F} f^+ d(\mu \times v) - \int_{E\times F} f^- d(\mu \times v) = \int_{E\times F} f d(\mu \times v).\end{aligned}$$

By a symmetric argument for the 1-section of f, we can see that f^y is μ-integrable for v-a.e. $y \in F$,

$$\psi(y) = \int_E f^y d\mu$$

is well-defined for v-a.e. $y \in E$, and that

$$\int_F \psi dv = \int_{E\times F} f d(\mu \times v) = \int_E \varphi d\mu.$$

Step 5: Complex-valued Functions

Let f be a complex valued function such that

$$\int_E \varphi^* d\mu < +\infty$$

for the function $\varphi^* : E \to [0, +\infty]$ defined as

$$\varphi^*(x) = \int_F |f|_x dv$$

for any $x \in E$. Then, because $|Re(f)| \leq |f|$, we can see that

$$\varphi^{**}(x) = \int_F |Re(f)|_x dv \leq \int_F |f|_x dv = \varphi^*(x)$$

for any $x \in E$, so that

$$\int_E \varphi^{**} d\mu \leq \int_E \varphi^* d\mu < +\infty.$$

By the preceding result, this implies that the $\mathcal{E} \bigotimes \mathcal{F}$-measurable real-valued function $Re(f)$ on $E \times F$ is $\mu \times v$-integrable.
By a symmetric argument, this holds for $Im(f)$ as well, so it follows that f is $\mu \times v$-integrable.

Now assume that f is a $\mu \times v$-integrable complex valued function. Then, by implication, $Re(f)$ and $Im(f)$ are $\mu \times v$-integrable real-valued functions, and by the preceding result, there exists an $A \in \mathcal{E}$ such that $Re(f)_x$ is v-integrable for any $x \in A$,

$$\varphi_1(x) = \int_F Re(f)_x dv$$

is well-defined for any $x \in A$, $\mu(A^c) = 0$, and

$$\int_{E\times F} Re(f) d(\mu \times v) = \int_E \varphi_1 d\mu.$$

Likewise, there exists a $B \in \mathcal{E}$ such that $Im(f)_x$ is v-integrable for any $x \in B$,

$$\varphi_2(x) = \int_F Im(f)_x dv$$

is well-defined for any $x \in B$, $\mu(B^c) = 0$, and

$$\int_{E \times F} Im(f) d(\mu \times v) = \int_E \varphi_2 d\mu.$$

Since $\mu((A \cap B)^c) = 0$, it follows that $f_x = Re(f)_x + i \cdot Im(f)_x$ is v-integrable for μ-a.e. $x \in E$,

$$\varphi(x) = \int_F f_x dv = \int_F Re(f)_x dv + i \cdot \int_F Im(f)_x dv = \varphi_1(x) + i \cdot \varphi_2(x)$$

is well-defined for the same $x \in E$, and

$$\begin{aligned}\int_{E \times F} f d(\mu \times v) &= \int_{E \times F} Re(f) d(\mu \times v) + i \cdot \int_{E \times F} Im(f) d(\mu \times v) \\ &= \int_E \varphi_1 d\mu + i \cdot \int_E \varphi_2 d\mu = \int_E \varphi d\mu.\end{aligned}$$

By a symmetric argument for the 1-section of f, we can seethat f^y is μ-integrable for v-a.e. $y \in F$,

$$\psi(y) = \int_E f^y d\mu$$

is well-defined for the same $y \in F$, and

$$\int_{E \times F} f d(\mu \times v) = \int_F \psi dv.$$

Q.E.D.

In the notation of the above theorem, we denote

$$\int_F f(x,y)dv(y) := \int_F f_x dv = \varphi(x) \quad \text{and} \quad \int_E f(x,y)d\mu(x) := \int_E f^y d\mu = \psi(y)$$

for any $x \in E$ and $y \in F$, and likewise,

$$\int_E \int_F f(x,y)dv(y)d\mu(x) := \int_E \varphi d\mu \quad \text{and} \quad \int_F int_E f(x,y)d\mu(x)dv(y) := \int_F \psi dv.$$

As such, the content of the theorem can be stated as

i) If $f \in (\mathcal{E} \bigotimes \mathcal{F})_+$, then

$$\int_F f(\cdot,y)dv(y) \in \mathcal{E}_+, \quad \int_E f(x,\cdot)d\mu(x) \in \mathcal{F}_+,$$

and

$$\int_{E\times F} fd(\mu \times v) = \int_E \int_F f(x,y)dv(y)d\mu(x) = \int_F \int_E f(x,y)d\mu(x)dv(y).$$

ii) If f is complex valued and

$$\int_E \int_F |f(x,y)|dv(y)d\mu(x) < +\infty,$$

then f is $\mu \times v$-integrable.

iii) If f is complex valued and $\mu \times v$-integrable, then f_x is v-integrable for μ-a.e. $x \in E$, f^y is μ-integrable for v-a.e. $y \in F$,

$$\int_E f(x,y)dv(y) \quad \text{and} \quad \int_F f(x,y)d\mu(x)$$

are well-defined for that same $x \in E$, $y \in F$, and

$$\int_{E\times F} fd(\mu \times v) = \int_E \int_F f(x,y)dv(y)d\mu(x) = \int_F \int_E f(x,y)d\mu(x)dv(y) \in \mathbb{C}.$$

In this sense, Fubini's theorem tells us that we can interchange the order of integration of an integrable function defined on a product space.

7.4 Transition Kernels and Product Spaces

Transition kernels provide us with a convenient way to construct probability measures on product spaces. In this section we study how to construct a probability measure on a product spcae given a transition probability kernel and a probability measure, and the reverse problem of deriving transition probability kernels and probability measures that constitute a given probability measure. The material in this section lays the groundwork for the study of conditional probabilities and densities in probability theory.

7.4.1 Construction of Probability Measures

Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces, and K a transition kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$. We saw in section 3.6 that, for any $f \in \mathcal{F}_+$, the non-negative function $T_K f$ on E defined as

$$(T_K f)(x) = \int_F f(y)K(x,dy)$$

for any $x \in E$ is non-negative $\mathcal{E}$-measurable, and that the operation T_K on $\mathcal{F}_+$ possess linearity and continuity properties.

Now let $f \in (\mathcal{E}\bigotimes\mathcal{F})_+$. Then, for any $x \in E$, the section f_x is $\mathcal{F}$-measurable, so that the integral

$$(T_K f)(x) = \int_F f_x(y)K(x,dy)$$

of f_x with respect to the measure $K(x,\cdot)$ is well-defined for any $x \in E$. We saw above that, if $K(x,\cdot)$ is σ-finite and does not depend on x, that is, if the integral above is with respect to a σ-finite measure, then the function $T_K f$ is $\mathcal{E}$-measurable.

We now show that $T_k f$ is $\mathcal{E}$-measurable, and that the operation T_k on $(\mathcal{E}\bigotimes\mathcal{F})_+$ possess the same linearity and continuity properties as the corresponding operation defined on $\mathcal{F}_+$, if K is a transition probability kernel.

The proof makes extensive use of the monotone class theorem for functions.

Lemma 7.6 Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces, and K a transition probability kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$. Then, for any $f \in (\mathcal{E}\bigotimes\mathcal{F})_+$, defining the function $T_K f : E \to [0,+\infty]$ as

$$(T_K f)(x) = \int_F f_x(y)K(x,dy)$$

for any $x \in E$, $T_K f \in \mathcal{E}_+$.
Furthermore, the operation $T_K : (\mathcal{E}\bigotimes\mathcal{F})_+ \to \mathcal{E}_+$ satisfies the following properties:

- **Linearity**:
  For any $a \in [0,+\infty)$ and $f,g \in (\mathcal{E}\bigotimes\mathcal{F})_+$,

$$T_K(af+g) = a\cdot T_K f + T_K g$$

- **Continuity under Increasing Limits**:
  For any increasing sequence $\{f_n\}_{n\in N_+} \subset (\mathcal{E}\bigotimes\mathcal{F})_+$ with pointwise limit f,

$$T_K f_n \nearrow T_K f.$$

Proof) We start in a more general setting in order to facilitate the proof.
For any bounded and real-valued or non-negative $\mathcal{E}\bigotimes\mathcal{F}$-measurable function f, we define

$$(T_K f)(x) = \int_F f_x(y)K(x,dy)$$

for any $x \in E$; since K is a transition probability kernel, $K(x,F) = 1$ for any $x \in E$, meaning that for any bounded real-valued $\mathcal{E}\bigotimes\mathcal{F}$-measurable function f, the above integral is well-defined.

Now let $\mathcal{M}$ be the collection of $\mathcal{E}\bigotimes\mathcal{F}$-measurable functions such that:

i) f is non-negative valued or bounded real-valued, so that $T_K f$ is well-defined
ii) $T_K f$ is $\mathcal{E}$-measurable.

Clearly, $\mathcal{M}$ contains the indicator functions of every measurable rectangle. To see this, let $A\times B$ be a measurable rectangle on $E\times F$, and note that

$$(T_K f)(x) = \int_F (I_{A\times B})_x(y)K(x,dy) = I_A(x)\cdot\int_F I_B(y)K(x,dy) = I_A(x)\cdot K(x,B)$$

for any $x \in E$. Since I_A and $K(\cdot,B)$ are non-negative measurable functions on E, it follows that $T_K f \in \mathcal{E}_+$ as well.

$\mathcal{M}$ is also a monotone class of functions on $E\times F$:

- $I_{E\times F} \in \mathcal{M}$ because $E\times F$ is a measurable rectangle.

- For any $a,b \in \mathbb{R}$ and bounded $f,g \in \mathcal{M}$, $af+bg$ is a $\mathcal{E}\bigotimes\mathcal{F}$-measurable bounded real-valued function.
  It also follows that

$$\begin{aligned}(T_K(af+bg))(x) &= \int_F (af+bg)_x(y)K(x,dy) \\ &= \int_F (a\cdot f_x(y)+b\cdot g_x(y))K(x,dy) = a\cdot\int_F f_x(y)K(x,dy)+b\dot{\int}_F g_x(y)K(x,dy) \\ &= a\cdot(T_K f)(x)+b\cdot(T_K g)(x)\end{aligned}$$

  for any $x\in E$ by the linearity of integration, since f_x, g_x are bounded and thus $K(x,\cdot)$-integrable for any $x\in E$. Since $T_K f, T_K g$ are $\mathcal{E}$-measurable by assumption, and measurability is preserved across linear combinations, $T_K(af+bg)$ is also $\mathcal{E}$-measurable, so that $af+bg\in\mathcal{M}$.

- For any increasing sequence $\{f_n\}_{n\in N_+}$ of non-negative functions in $\mathcal{M}$ with pointwise limit f, $f\in(\mathcal{E}\bigotimes\mathcal{F})_+$ because measurability is preserved across limits.
  Furthermore, for any $x\in E$, because $\{(f_n)_x\}_{n\in N_+}$ is an increasing sequence of non-negative $\mathcal{F}$-measurable functions with pointwise limit f_x for any $x\in E$, by the MCT we have

$$\begin{aligned}(T_K f)(x) = \int_F f_x(y)K(x,dy) &= \lim_{n\to\infty}\int_F (f_n)_x(y)K(x,dy) \\ &= \lim_{n\to\infty}(T_K f_n)(x).\end{aligned}$$

  By the monotonicity of integration and the fact that $(f_n)_x \le (f_{n+1})_x$ for any $n\in N_+$, we also have $(T_K f_n)(x)\le(T_K f_{n+1})(x)$ for any $n\in N_+$.
  Therefore, $T_K f$ is the pointwise limit of an increasing sequence $\{T_K f_n\}_{n\in N_+}$ of non-negative $\mathcal{E}$-measurable functions, so that $T_K f\in\mathcal{E}_+$ and therefore $f\in\mathcal{M}$.

$\mathcal{M}$ is thus a monotone class of functions on $E\times F$ containing every measurable rectangle on $E\times F$. Since the collection of all measurable rectangles on $E\times F$ is a π-system generating the product σ-algebra $\mathcal{E}\bigotimes\mathcal{F}$, by the monotone class theorem for functions, it follows that every bounded or non-negative $\mathcal{E}\bigotimes\mathcal{F}$-measurable function is contained in $\mathcal{M}$. By implication, $T_K f\in\mathcal{E}_+$ for any $f\in(\mathcal{E}\bigotimes\mathcal{F})_+$.

To show that T_K possess the linearity and continuity properties stated above, we rely on the fact that $\mathcal{M}$ is a montone class of functions. We already proved, in the process of showing that $\mathcal{M}$ is a monotone class, that for any increasing sequence $\{f_n\}_{n\in N_+}$ of non-negative functions in $\mathcal{M}$ with pointwise limit f,

$$T_K f_n \nearrow T_K f.$$

Thus, it remains to show that T_K is linear. To this end, choose any $a \in [0,+\infty)$ and $f,g \in (\mathcal{E} \bigotimes \mathcal{F})_+$. Letting $\{f_n\}_{n\in N_+}$, $\{g_n\}_{n\in N_+}$ be sequences of measurable simple functions increasing to f and g, because f_n, g_n are bounded for any $n \in N_+$, from the result shown above while proving that $\mathcal{M}$ satisfies the second property of monotone classes,

$$T_K(af_n + g_n) = a \cdot T_K f_n + T_K g_n.$$

Since $\{af_n + g_n\}_{n\in N_+}$ is an increasing sequence of measurable functions with limit $af + g$, it now follows from the continuity of T_K under increasing limits that

$$\begin{aligned} T_K(af+g) &= \lim_{n\to\infty} T_K(af_n + g_n) \\ &= a \cdot \left(\lim_{n\to\infty} T_K f_n\right) + \lim_{n\to\infty} T_K g_n = a \cdot T_K f + T_K g. \end{aligned}$$

Q.E.D.

Let μ be a measure on $(E,\mathcal{E})$. In light of the above result, it makes sense to define a function $\pi : \mathcal{E} \bigotimes \mathcal{F} \to [0,+\infty]$ as

$$\pi(A) = \int_E (T_K I_A) d\mu := \int_E \int_F I_A(x,y) K(x,dy) d\mu(x)$$

for any $A \in \mathcal{E} \bigotimes \mathcal{F}$, like we did when constructing product measures. The next result shows that, when μ is taken to be a probability measure, the π defined above becomes a probability measure on the product space.

Theorem 7.7 (Construction of Probability Measures)
Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces, and K a transition probability kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$. For any probability measure μ on $(E,\mathcal{E})$, defining the function $\pi : \mathcal{E}\bigotimes\mathcal{F} \to [0,+\infty]$ as

$$\pi(A) = \int_E (T_K I_A) d\mu$$

for any $A \in \mathcal{E}\bigotimes\mathcal{F}$, π is a probability measure on $(E\times F, \mathcal{E}\bigotimes\mathcal{F})$, and for any $f \in (\mathcal{E}\bigotimes\mathcal{F})_+$,

$$\int_{E\times F} f d\pi = \int_E (T_K f) d\mu.$$

Furthermore, π is the unqiue probability measure on $(E\times F, \mathcal{E}\bigotimes\mathcal{F})$ such that

$$\pi(A\times B) = \int_A K(x,B) d\mu(x)$$

for any measurable rectangle $A\times B$ on $E\times F$.

Proof) We again rely on the characterization of integration introduced in theorem 3.16. Define the function $\Lambda : (\mathcal{E}\bigotimes\mathcal{F})_+ \to [0,+\infty]$ as

$$\Lambda f = \int_E (T_K f) d\mu$$

for any $f \in (\mathcal{E}\bigotimes\mathcal{F})_+$. For any measurable rectangle $A\times B$ on $E\times F$, because

$$T_K I_{A\times B} = K(\cdot,B)\cdot I_A,$$

we have

$$\Lambda I_{A\times B} = \int_A K(\cdot,B) d\mu.$$

We will now show that Λ satisfies the conditions of theorem 3.16:

- $\Lambda I_\emptyset = 0$ is clear.

- For any $a \in [0,+\infty)$ and $f,g \in (\mathcal{E}\bigotimes\mathcal{F})_+$, we have

  $$T_K(af+g) = a\cdot T_K f + T_K g$$

  by the linearity of T_K. It then follows from the linearity of integration that

  $$\Lambda(af+g) = \int_E (a\cdot T_K f + T_K g)\, d\mu = a\cdot \int_E (T_K f) d\mu + \int_E (T_K g) d\mu = a\cdot\Lambda f + \Lambda g.$$

- For any increasing sequence of non-negative $\mathcal{E}\bigotimes\mathcal{F}$-measurable functions $\{f_n\}_{n\in N_+}$

with pointwise limit f, by the continuity of T_K under increasing limits,

$$T_K f_n \nearrow T_K f.$$

$\{T_K f_n\}_{n \in N_+}$ is thus a sequence of $\mathcal{E}$-measurable non-negative functions increasing to $T_K f$, so by the MCT,

$$\begin{aligned}\Lambda f = \int_E (T_K f) d\mu &= \lim_{n\to\infty} \int_E (T_K f_n) d\mu \\ &= \lim_{n\to\infty} \Lambda f_n.\end{aligned}$$

Furthermore, by the monotonicity of integration, $\Lambda f_n \leq \Lambda f_{n+1}$ for any $n \in N_+$, so that $\Lambda f_n \nearrow \Lambda f$.

Therefore, by theorem 3.16, there exists a unique measure π' on $(E \times F, \mathcal{E} \bigotimes \mathcal{F})$ such that

$$\Lambda f = \int_{E\times F} f d\pi'$$

for any $f \in (\mathcal{E} \bigotimes \mathcal{F})_+$. For any $A \in \mathcal{E} \bigotimes \mathcal{F}$,

$$\begin{aligned}\pi(A) = \int_E (T_K I_A) d\mu &= \Lambda I_A \\ &= \int_{E\times F} I_A d\pi' = \pi'(A),\end{aligned}$$

so that $\pi = \pi'$ on $\mathcal{E} \bigotimes \mathcal{F}$ and π is a measure on the product space $(E \times F, \mathcal{E} \bigotimes \mathcal{F})$. Since

$$\pi(E \times F) = \Lambda I_{E\times F} = \int_E K(x, F) d\mu(x) = \mu(E) = 1$$

by the fact that K is a transition probability kernel and μ a probability measure, π is also a probability measure.

Finally, suppose that $\overline{\pi}$ is another probability measure on $(E \times F, \mathcal{E} \bigotimes \mathcal{F})$ such that

$$\overline{\pi}(A \times B) = \int_A K(x, B) d\mu(x)$$

for any measurable rectangle $A \times B$ on $E \times F$. Then, for any measurable rectangle $A \times B$,

$$\pi(A \times B) = \Lambda I_{A\times B} = \int_A K(x, B) d\mu(x) = \overline{\pi}(A \times B),$$

so that $\pi = \overline{\pi}$ on the set $\mathcal{R}$ of all measurable rectangles on $E \times F$. Since $\mathcal{R}$ is a π-system generating the product σ-algebra $\mathcal{E}\bigotimes\mathcal{F}$, and $\pi,\overline{\pi}$ are finite measures such that

$$\pi(E\times F)=\overline{\pi}(E\times F)=1<+\infty,$$

by lemma 2.15 $\pi=\overline{\pi}$ on $\mathcal{E}\bigotimes\mathcal{F}$, which proves uniqueness.
Q.E.D.

We denote the value $\int_E(T_K f)d\mu$ by

$$\int_E\int_F f(x,y)K(x,dy)d\mu(x)$$

for any $\mathcal{E}\bigotimes\mathcal{F}$-measurable non-negative function f, so that the content of the above theorem is that

$$\int_{E\times F} f d\pi = \int_E\int_F f(x,y)K(x,dy)d\mu(x)$$

for any $f\in(\mathcal{E}\bigotimes\mathcal{F})_+$. This relationship is succinctly referred to as

$$\pi=\mu\times K.$$

We have thus constructed a probability measure on the product space by use of a probability measure on $(E,\mathcal{E})$ and a transition probability kernel relating $(E,\mathcal{E})$ with $(F,\mathcal{F})$. A product probability measure is a special case of the above construction in which $K(x,\cdot)$ does not depend on x. This means that $(E,\mathcal{E})$ and $(F,\mathcal{F})$ are not linked together and, in a sense, independent of one another. It is precisely this mathematical intuition that informs the measure-theoretic definition of the independence of random variables.

7.4.2 The Density of a Transition Kernel

Fubini's theorem and the product space operations we have studied so far can be used to define something akin to the Radon-Nikodym derivative for transition kernels.

Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, v)$ be σ-finite measure spaces, and $k \in (\mathcal{E} \bigotimes \mathcal{F})_+$. Define the function $K : E \times \mathcal{F} \to [0, +\infty]$ as

$$K(x, A) = \int_A k(x, y) dv(y) = \int_F (k_x \cdot I_A) \, dv$$

for any $(x, A) \in E \times \mathcal{F}$; K is well-defined because the section k_x is non-negative $\mathcal{F}$-measurable, as is the indicator I_A.
We can now show that K is a transition kernel:

- For any $A \in \mathcal{F}$, define the function $k_A : E \times F \to [0, +\infty]$ as

  $$k_A(x, y) = k(x, y) \cdot I_A(y)$$

  for any $(x, y) \in E \times F$. Since k_A is the product of k and $I_{E \times A}$, which are both $\mathcal{E} \bigotimes \mathcal{F}$-measurable non-negative functions, $k_A \in (\mathcal{E} \bigotimes \mathcal{F})_+$ as well. By implication,

  $$K(\cdot, A) = \int_A k(\cdot, y) dv(y) = \int_F k_A(\cdot, y) dv(y)$$

  is a $\mathcal{E}$-measurable non-negative function by Fubini's theorem.

- For any $x \in E$,

  $$K(x, A) = \int_A k(x, y) dv(y) = \int_A k_x dv.$$

  for any $A \in \mathcal{F}$. Since $k_x \in \mathcal{F}_+$, this simply means that $K(x, \cdot)$ is the indefinite integral of k_x with respect to the measure v; therefore, $K(x, \cdot)$ is a measure on $(F, \mathcal{F})$, and k_x is its Radon-Nikodym derivative with respect to the σ-finite measure v.

Note that, if $\int_F k(x, y) dv(y) = 1$, then $K(x, F) = 1$ and K is a transition probability kernel.
Due to the fact that

$$k_x \in \frac{\partial K(x, \cdot)}{\partial v}$$

for any $x \in E$ in the above setting, we call k the density of the transition kernel K.
As with the usual Radon-Nikodym derivatives, $K(x, \cdot)$ is absolutely continuous with respect to v. That is, if $v(A) = 0$ for some $A \in \mathcal{F}$, then $K(x, A) = 0$ for any $x \in E$.

7.4.3 Disintegration of Probability Measures and Bayes' Rule

Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},v)$ be σ-finite measure spaces. Recall that the product measure $\mu\times v$ is also a σ-finite measure.

Suppose we have a probability measure π on the product space $(E\times F,\mathcal{E}\bigotimes\mathcal{F})$, and suppose that π is absolutely continuous with respect to the product measure $\mu\times v$, so that there exists a density function $f\in(\mathcal{E}\bigotimes\mathcal{F})_+$ such that

$$\pi(A)=\int_A fd(\mu\times v)=\int_E\int_F f(x,y)I_A(x,y)dv(y)d\mu(x)$$

for any $A\in\mathcal{E}\bigotimes\mathcal{F}$.

This section deals with how to construct a probability measure P on $(E,\mathcal{E})$ and a transition probability kernel K from $(E,\mathcal{E})$ to $(F,\mathcal{F})$ such that $\pi=P\times K$ given that π has a density f with respect to the product measure $\mu\times v$. This process is referred to as the disintegration of π.

Some preliminary results are in order.

Let $(G,\mathcal{G})$ be a measurable space and $f:E\to G$ a function measurable relative to $\mathcal{E}$ and $\mathcal{G}$. Defining $f_e:E\times F\to G$ as

$$f_e(x,y)=f(x)$$

for any $(x,y)\in E\times F$, for any $A\in\mathcal{G}$ we have

$$f_e^{-1}(A)=f^{-1}(A)\times F\in\mathcal{E}\bigotimes\mathcal{F}$$

since $f^{-1}(A)\in\mathcal{E}$, so that f_e is measurable relative to $\mathcal{E}\bigotimes\mathcal{F}$ and $\mathcal{G}$. The subscript e will thus denote the extension of the domain of functions defined only on an individual space to the product space.

Furthermore, for notational clarity, we will adopt the convention introduced in the previous sections that

$$\varphi(x)=\int_F f_x dv \quad \text{and} \quad \psi(y)=\int_E f^y d\mu$$

for any $x\in E$, $y\in F$. Fubini's theorem showed us that φ and ψ are measurable functions.

For any $B\in\mathcal{F}$, because

$$\int_B f_x dv=\int_F(f\cdot I_{E\times B})_x dv,$$

where $f\cdot I_{E\times B}$ is $\mathcal{E}\bigotimes\mathcal{F}$-measurable, the function $\varphi_B:E\to[0,+\infty]$ defined as

$$\varphi_B(x)=\int_B f_x dv$$

for any $x\in E$ is also $\mathcal{E}$-measurable. We adopt this notation in the proof of the next theorem:

Theorem 7.8 (Disintegration of Probability Measures)
Let $(E,\mathcal{E},\mu)$ and $(F,\mathcal{F},v)$ be σ-finite measure spaces, and π a probability measure on the product space $(E\times F,\mathcal{E}\bigotimes\mathcal{F})$ that has density $f\in(\mathcal{E}\bigotimes\mathcal{F})_+$ with respect to the product measure $\mu\times v$. Then, the following hold true:

i) Defining $P:\mathcal{E}\to[0,1]$ as

$$P(A)=\pi(A\times F)$$

for any $A\in\mathcal{E}$, P is a probability measure on $(E,\mathcal{E})$, and the non-negative function f_X on E defined as

$$f_X(x)=\int_F f(x,y)dv(y)$$

for any $x\in E$ is a density of P with respect to the σ-finite measure μ.

ii) Defining $k:E\times F\to[0,+\infty]$ as

$$k(x,y)=\begin{cases}\frac{f(x,y)}{f_X(x)} & \text{if } f_X(x)\in(0,+\infty)\\ \int_E f(x,y)d\mu(x) & \text{otherwise}\end{cases}$$

for any $(x,y)\in E\times F$, k is a non-negative $\mathcal{E}\bigotimes\mathcal{F}$-measurable function such that

$$f(x,y)=f_X(x)k(x,y)$$

for any $y\in F$ and P-a.e. $x\in E$. Moreover,

$$\pi(A)=\int_E\int_F f_X(x)k(x,y)I_A(x,y)dv(y)d\mu(x).$$

for any $A\in\mathcal{E}\bigotimes\mathcal{F}$.

iii) Defining $K:E\times\mathcal{F}\to[0,1]$ as

$$K(x,A)=\int_A k(x,y)dv(y)$$

for any $(x,A)\in E\times\mathcal{F}$, K is a transition probability kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$ such that

$$\pi=P\times K.$$

Proof) We again proceed in steps:

Step 1: Defining P and f_X

Define $P : \mathcal{E} \to [0,1]$ as

$$P(A) = \pi(A \times F)$$

for any $A \in \mathcal{E}$. Defining $f_X : E \to [0,+\infty]$ as

$$f_X(x) = \int_F f(x,y) dv(y) = \int_F f_x dv$$

for any $x \in E$, we know that $f_X \in \mathcal{E}_+$ from Fubini's theorem, and Fubini's theorem also tells us that

$$P(A) = \pi(A \times F) = \int_A f_X d\mu$$

for any $A \in \mathcal{E}$. This means that P is a measure on $(E, \mathcal{E})$ with density f_X with respect to μ. Since

$$P(E) = \pi(E \times F) = 1$$

because π is a probability measure, it follows that P is also a probability measure. This proves the first claim of the theorem.

Step 2: Defining and Proving the Measurability of k

Now define $k : E \times F \to [0,+\infty]$ as

$$k(x,y) = \begin{cases} \frac{f(x,y)}{f_X(x)} & \text{if } f_X(x) \in (0,+\infty) \\ \int_E f(x,y)d\mu(x) & \text{otherwise} \end{cases}$$

for any $(x,y) \in E \times F$. Define the sets $A_\infty, A_0 \in \mathcal{E}$ as

$$A_\infty = \{f_X = +\infty\} \quad \text{and} \quad A_0 = \{f_X = 0\}.$$

Since

$$\int_E f_X d\mu = P(E) = 1 < +\infty,$$

by the finiteness property of non-negative functions $\mu(A_\infty) = 0$; by the absolute continuity of P with respect to μ, this implies that $P(A_\infty) = 0$.
On the other hand,

$$P(A_0) = \int_{\{f_X = 0\}} f_X d\mu = 0,$$

so that, denoting $N = A_0 \cap A_\infty \in \mathcal{E}$, we have

$$P(N) = P(A_0) + P(A_\infty) = 0.$$

For any $x \in N^c$ and $y \in F$, $f_X(x) \in (0,+\infty)$ and thus

$$f(x,y) = k(x,y) f_X(x)$$

by the definition of k. In other words, the above equality holds for any $y \in F$ and P-a.e. $x \in E$.

For any $a \in \mathbb{Q}$ and $(x,y) \in E \times F$, if $x \in N^c$, then

$$k(x,y) < a \text{ if and only if } f(x,y) < a \cdot f_X(x) = a \cdot f_{X,e}(x,y),$$

where $f_{X,e}$ is f_X with its domain extended to $E \times F$.
On the other hand, if $x \in N$, then

$$k(x,y) < a \text{ if and only if } \psi(y) < a.$$

Therefore,

$$\{(x,y)\in E\times F \mid k(x,y)<a\} = [\{(x,y)\in E\times F \mid f(x,y)<af_{X,e}(x,y)\}\cap(N^c\times F)] \\ \cup[N\times\{y\in F \mid \psi(y)<a\}].$$

By the $\mathcal{F}$-measurability of ψ, $\{y\in F \mid \psi(y)<a\}\in\mathcal{F}$, and by the measurability of f and $a\cdot f_{X,e}$ on $\mathcal{E}\bigotimes\mathcal{F}$, $\{(x,y)\in E\times F \mid f(x,y)<af_{X,e}(x,y)\}\in\mathcal{E}\bigotimes\mathcal{F}$. It follows that

$$\{(x,y)\in E\times F \mid k(x,y)<a\}\in\mathcal{E}\bigotimes\mathcal{F}.$$

and because this holds for any $a\in\mathbb{Q}$, k is $\mathcal{E}\bigotimes\mathcal{F}$-measurable.

Step 3: Expressing π in terms of k and f_X

For any measurable rectangle $A \times B$ on $E \times F$,

$$\begin{aligned}\pi(A\times B) &= \int_E \int_F f(x,y) I_{A\times B}(x,y) dv(y) d\mu(x) \\ &= \int_A \varphi_B d\mu \\ &= \int_{A\cap N} \varphi_B d\mu + \int_{A\cap N^c} \varphi_B d\mu \\ &= \pi((A\cap N)\times B) + \int_{A\cap N^c} \varphi_B d\mu.\end{aligned}$$

We inspect each term in turn.

Because $f_X(x) = 0$ for any $x \in A_0$,

$$\pi((A\cap A_0)\times B) \leq \pi((A\cap A_0)\times F) = P(A\cap A_0) = \int_{A\cap A_0} f_X(x) d\mu(x) = 0,$$

where we used the fact that f_X is the density of P with respect to μ.
Likewise, since $\mu(A_\infty) = 0$,

$$\begin{aligned}\pi((A\cap A_\infty)\times B) \leq \pi((A\cap A_\infty)\times F) &= P(A\cap A_\infty) \\ &= \int_{A\cap A_\infty} f_X(x) d\mu(x) = 0.\end{aligned}$$

These inequalities imply that $\pi((A\cap A_0)\times B) = \pi((A\cap A_\infty)\times B) = 0$, and as such that

$$\pi((A\cap N)\times B) = \pi((A\cap A_0)\times B) + \pi((A\cap A_\infty)\times B) = 0.$$

Now we turn our attention to the second term.
Define $g : E \times F \to [0, +\infty]$ as $g = k \cdot f_{X,e}$. Because both k and $f_{X,e}$ are measurable, $g \in (\mathcal{E} \bigotimes \mathcal{F})_+$.
If $x \in N^c$, then $f_X(x) \in (0, +\infty)$ and $g(x,y) = k(x,y) f_X(x) = f(x,y)$, so that

$$\varphi_B(x) = \int_B f_x dv = \int_B g_x dv$$

for any $x \in A \cap N^c$. Defining $\phi_B : E \to [0, +\infty]$ as

$$\phi_B(x) = \int_B g_x dv$$

for any $x \in E$, ϕ_B is $\mathcal{E}$-measurable by the same reason φ_B is, and it follows that

$$\int_{A\cap N^c} \varphi_B d\mu = \int_{A\cap N^c} \phi_B d\mu.$$

Finally, note that if $x \in A_0$, then $f_X(x) = 0$, so that $g(x,y) = 0$ for any $(x,y) \in A_0 \times F$ and

$$\int_{A \cap A_0} \phi_B d\mu = 0,$$

while

$$\int_{A \cap A_\infty} \phi_B d\mu = 0$$

because $\mu(A_\infty) = 0$. By implication,

$$\int_A \phi_B d\mu = \int_{A \cap N^c} \phi_B d\mu + \int_{A \cap A_0} \phi_B d\mu + \int_{A \cap A_\infty} \phi_B d\mu = \int_{A \cap N^c} \phi_B d\mu.$$

Therefore,

$$\begin{aligned}\pi(A \times B) &= \pi((A \cap N) \times B) + \int_{A \cap N^c} \varphi_B d\mu = \int_A \phi_B d\mu \\ &= \int_E \int_F g(x,y) I_{A \times B}(x,y) dv(y) d\mu(y).\end{aligned}$$

Defining $\overline{\pi} : \mathcal{E} \bigotimes \mathcal{F} \to [0, \infty]$ as

$$\overline{\pi}(A) = \int_A g d(\mu \times v) = \int_E \int_F g(x,y) I_A(x,y) dv(y) d\mu(x)$$

for any $A \in \mathcal{E} \bigotimes \mathcal{F}$, $\overline{\pi}$ is a measure on the product space because g is measurable and non-negative.

We have shown above that

$$\pi(A \times B) = \int_E \int_F g(x,y) I_{A \times B}(x,y) dv(y) d\mu(y) = \overline{\pi}(A \times B)$$

for any measurable rectangle $A \times B$ on $E \times F$. Since $\mathcal{E} \bigotimes \mathcal{F}$ is generated by the π-sysetm of measurable rectangles on $E \times F$ and

$$\overline{\pi}(E \times F) = \pi(E \times F) = 1 < +\infty,$$

by lemma 2.15 $\pi = \overline{\pi}$ on $\mathcal{E} \bigotimes \mathcal{F}$. In other words, for any $A \in \mathcal{E} \bigotimes \mathcal{F}$,

$$\pi(A) = \int_A g d(\mu \times v) = \int_E \int_F f_X(x) k(x,y) I_A(x,y) dv(y) d\mu(x).$$

Step 4: Showing that $\pi = P \times K$

Lastly, define the function $K : E \times \mathcal{F} \to [0,+\infty]$ as

$$K(x,A) = \int_A k(x,y)dv(y)$$

for any $(x,A) \in E \times \mathcal{F}$. We now know that K is a transition kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$ with k as its density with respect to v.
In addition, if $x \in N^c$, then we have

$$K(x,F) = \int_F k(x,y)dv(y) = \frac{\int_F f(x,y)dv(y)}{f_X(x)} = 1$$

by the definition of $f_X(x)$, and if $x \in N$, then

$$\begin{aligned} K(x,F) &= \int_F k(x,y)dv(y) = \int_F \int_E f(x,y)d\mu(x)dv(y) \\ &= \int_{E\times F} f d(\mu \times v) = \pi(E \times F) = 1. \end{aligned}$$

Therefore, K is a transition probability kernel from $(E,\mathcal{E})$ into $(F,\mathcal{F})$.

From the result we showed above, for any measurable rectangle $A \times B$ on $E \times F$,

$$\begin{aligned} \pi(A \times B) &= \int_E \int_F f_X(x)k(x,y)I_{A\times B}(x,y)dv(y)d\mu(x) \\ &= \int_E f_X(x)K(x,B)I_A(x)d\mu(x). \end{aligned}$$

Because f_X is the density of P with respect to μ and $K(\cdot,B)I_A$ is in $\mathcal{E}_+$,

$$\int_E f_X(x)K(x,B)I_A(x)d\mu(x) = \int_E K(x,B)I_A(x)dP(x) = \int_A K(x,B)dP(x)$$

and therefore

$$\pi(A \times B) = \int_A K(x,B)dP(x).$$

From theorem 7.7, we know that $P \times K$ is the unique probability measure on the product space such that

$$(P \times K)(A \times B) = \int_A K(x,B)dP(x)$$

for any measurable rectangle $A \times B$; therefore,

$$\pi = P \times K$$

on the product space.
Q.E.D.

In probabilistic terms, if π is the joint distribution of two random variables X and Y, the probability measure P defined above as $P(A) = \pi(A \times F)$ for any $A \in \mathcal{E}$ is the marginal distribution of X. This would make the transition probability kernel K the conditional distribution of Y given X.
The density f_X is then the marginal density of X with respect to the measure μ, and $k(x, \cdot)$ is the conditional density of Y given $X = x$ for any $x \in E$ with respect to the measure v.
We also showed that

$$f(x,y) = k(x,y) f_X(x), \quad f_X(x) \in (0, +\infty)$$

for any $y \in F$ and P-a.e. $x \in E$, which also implies that this holds for π-a.e. $(x,y) \in E \times F$.

While the decomposition above was performed from the perspective of the first variable X, it can also be done for the second variable Y. In this case, we can define the function $f_Y : F \to [0, +\infty]$ as

$$f_Y(y) = \int_E f(x,y) d\mu(y)$$

for any $y \in F$ and the function $q : E \times F \to [0, +\infty]$ as

$$q(y,x) = \begin{cases} \frac{f(x,y)}{f_Y(y)} & \text{if } f_Y(y) \in (0, +\infty) \\ \int_F f(x,y) dv(y) & \text{othewise} \end{cases}$$

for any $(x,y) \in E \times F$.
As in the decomposition from the perspective of X, f_Y is the marginal density of Y with respect to v, and $q(y, \cdot)$ the conditional density of X given $Y = y$ for any $y \in F$ with respect to μ. It also follows that

$$f(x,y) = q(y,x) f_Y(y), \quad f_Y(y) \in (0, +\infty)$$

for π-a.e. $(x,y) \in E \times F$.

Bringing the two results together, we can see that

$$k(x,y) f_X(x) = f(x,y) = q(y,x) f_Y(y) \quad \text{and} \quad f_X(x), f_Y(y) \in (0, +\infty)$$

for $\mu \times v$-a.e. $(x,y) \in E \times F$. By implication,

$$k(x,y) = \frac{q(y,x) f_Y(y)}{f_X(x)}$$

for π-a.e. $(x,y) \in E \times F$, which is precisely the content of Bayes' rule.