Probability and Measure Notes

Alexander Yuan

September 22 2025

1 Measure Theory

1.1 Sigma Algebras

Suppose Ω is a non-empty set and a collection $\mathcal{F} \subset \mathcal{P}(\Omega)$ satisfy:

1) $\Omega \in \mathcal{F}$

2) $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F} \implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$

3) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$

Such a collection $\mathcal{F}$ is called a σ-algebra over Ω. A collection that satisfies 1) and 3) but is only closed under finite unions is called an *algebra*. Some authors also use field instead of algebra.

A sigma algebra is closed under countably infinite intersections and an algebra is closed under finite intersections. This follows immediately from *DeMorgan's law(s)*.

The *σ-algebra generated by a class* $\mathcal{C} \subset \mathcal{P}(\Omega)$ denoted $\sigma(\mathcal{C})$, is referred to as the smallest sigma algebra containing $\mathcal{C}$ and is defined by taking the intersection of all sigma algebras that contain $\mathcal{C}$.

1.1.1 Toy σ-algebra:

Let $\Omega = \{a, b, c, d\}$, and consider the classes

$$\mathcal{F}_1 = \{\Omega, \emptyset, \{a\}\}$$

$$\mathcal{F}_2 = \{\Omega, \emptyset, \{a\}, \{b, c, d\}\}$$

$\mathcal{F}_2$ is a σ-algebra (and an algebra), but $\mathcal{F}_1$ is neither.

1.1.2 Finite/Cofinite σ-algebra

Let Ω be a nonempty set, and let $|A|$ denote the number of elements of a set $A \subset \Omega$. Define the collection $\mathcal{F}_3 = \{A \subset \Omega : \text{ either } |A| \text{ is finite or } |A^c| \text{ is finite.}\}$ $\mathcal{F}_3$ is a σ-algebra if and only if $|\Omega| < \infty$.

1.1.3 Lemma: If $\mathcal{C} \subset \sigma(\mathcal{D})$ then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$

Proof.

$\sigma(\mathcal{D})$ is a sigma algebra that contains $\mathcal{C}$, it therefore also contains $\sigma(\mathcal{C})$.

1.1.4 Trivial σ-algebra(s)

For any set Ω, the collection $\{\emptyset, \Omega\}$, as well as the power set $\mathcal{P}(\Omega)$ is a sigma algebra. We can think of these respectively as the "smallest" and "largest" sigma algebras we can define on Ω.

1.1.5 Semi-Algebra

Another collection of subsets of interest is the *semi-algebra*. Let Ω be a non-empty set. A collection $\mathcal{C} \subset \mathcal{P}(\Omega)$ is called a semi-algebra if:

$$1) \ A, B \in \mathcal{C} \implies A \cap B \in \mathcal{C}$$

$$2) \text{ For any } A \in \mathcal{C}, \text{ there exist sets } B_1, \ldots, B_k \in \mathcal{C} \text{ such that } A^c = \bigcup_{i=1}^{k} B_i$$

1.1.6 Half-open/ray Semi-algebra

Let $\Omega = \mathbb{R}$ and $\mathcal{C} := \{(a, b], (b, \infty) : -\infty \leq a, b < \infty\}$, then $\mathcal{C}$ is a semi-algebra.

1.1.7 Interval Semi-algebra

Define an interval in $\mathbb{R}$ as a set $I \subset \mathbb{R}$ such that $a, b \in I, a < b \implies (a, b) \subset I$. Let $\Omega = \mathbb{R}$ and $\mathcal{C} := \{I : I \text{ is an interval}\}$, then $\mathcal{C}$ is a semi-algebra.

1.1.8 Borel σ-algebra

If Ω is a metric space (or more generally any topological space), then the σ-algebra generated by the topology of Ω is called the Borel σ-algebra on Ω. It contains all open sets and closed subsets of Ω, all countable intersections of open sets, and all countable unions of closed sets.

We denote the Borel σ-algebra on $\mathbb{R}$ as $\mathcal{B}_{\mathbb{R}}$. For example, if Ω is Hausdorff, then $\mathcal{B}_\Omega$ contains all countable and co-countable sets.

1.1.9 Borel σ-algebra generation

The open intervals, closed intervals, half open intervals, open and closed rays all generate $\mathcal{B}_{\mathbb{R}}$. When we construct the Borel Measure, it will be the half-open intervals of most interest.

Proof.

(Verify this!).

1.1.10 Product σ-algebra

Let $\{X_\alpha\}_{\alpha \in A}$ be an indexed collection of non empty sets and $\Omega = \prod_{\alpha \in A} X_\alpha$ and coordinate maps $\pi_\alpha : \Omega \to X_\alpha$. If $\mathcal{F}_\alpha$ is a σ-algebra on X_α for each α, the *product σ-algebra* on Ω is the σ-algebra generated by:

$$\{\pi^{-1}(E_\alpha) : E_\alpha \in \mathcal{F}_\alpha,\ \alpha \in A\}$$

We denote this by $\bigotimes_{\alpha \in A} \mathcal{F}_\alpha$. This definition is a little abstract, keep in mind it is incredibly significant for maps between measurable spaces.

1.1.11 Countably generated σ-algebra

If the index set A for $\{X_\alpha\}_{\alpha \in A}$ is countable, then $\bigotimes_{\alpha \in A} \mathcal{F}_\alpha$ is the σ-algebra generated by $\{\prod_{\alpha \in A} E_\alpha : E_\alpha \in \mathcal{F}_\alpha\}$.

Proof.

If $E_\alpha \in \mathcal{F}_\alpha$, then $\pi_\alpha^{-1}(E_\alpha) = \prod_{\beta \in A} E_\beta$ where $E_\beta = X_\beta$ for $\beta \neq \alpha$. Also note $\prod_{\alpha \in A} E_\alpha = \bigcap_{\alpha \in A} \pi_\alpha^{-1}$ and the result follows from (1.1.1).

1.1.12 Product (Cylindrical) σ-algebra generation

Suppose that $\mathcal{F}_\alpha$ is generated by a class $\mathcal{E}_\alpha$, $\alpha \in A$. Then $\bigotimes_{\alpha \in A} \mathcal{F}_\alpha$ is generated by $\mathcal{J}_1 := \{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E},\ \alpha \in A\}$. If the index set A is countable and $X_\alpha \in \mathcal{E}_\alpha$ for all α, $\bigotimes_{\alpha \in A} \mathcal{F}_\alpha$ is generated by $\mathcal{J}_2 := \{\prod_{\alpha \in A} E_\alpha : E_\alpha \in \mathcal{E}_\alpha\}$.

Proof.

(Verify This!)

Note: The second assertion follows from the previous proposition.

1.1.13 Product σ-algebra reduction

Let $\Omega_1, \ldots, \Omega_n$ be metric spaces and let $\Omega = \prod_{i=1}^n \Omega_i$, equipped with the metric on the cartesian product of finitely many metric spaces. Note that this product metric d on $\Omega_1 \times \Omega_2$ is defined by $d((x_1, x_2), (y_1, y_2)) = \max\{d(x_1, y_1), d_2(x_2, y_2)\}$

Then $\bigotimes_{i=1}^{n} \mathcal{B}_{\Omega_i} \subset \mathcal{B}_\Omega$. Recall a metric space Ω_i is separable provided there is a coutable dense subset in Ω_i. If the Ω_i's are separable, then $\bigotimes_{i=1}^{n} \mathcal{B}_{\Omega_i} = \mathcal{B}_\Omega$.

Proof.

By the previous proposition, $\bigotimes_{i=1}^{n} \mathcal{B}_{\Omega_i}$ is generated by the sets $\pi_i^{-1}(U_i)$, for $1 \leq i \leq n$ where U_i is open in Ω_i. Since these sets are open in Ω, it follows that $\bigotimes_{i=1}^{n} \mathcal{B}_{\Omega_i} \subset \mathcal{B}_\Omega$ by (1.1.1). Suppose that C_i is a countable dense set in Ω, and let $\mathbb{B}_i$ be the collection of (open) balls in Ω_i with rational radius and center in C_i. Then every open set in Ω_i is a countable union of members of $\mathbb{B}_i$. Moreover, the set of points in Ω whose ith coordinate is in C_i for all i is a countable dense subset of Ω, and the balls of radius r in Ω turn out to be products of balls of radius r in the Ω_i's. It follows that $\mathcal{B}_{\Omega_i}$ is generated by $\mathbb{B}_i$ and $\mathcal{B}_\Omega$ is generated by $\{\prod_{i=1}^{n} B_i : B_i \in \mathbb{B}_i\}$. Therefore, $\bigotimes_{i=1}^{n} \mathcal{B}_{\Omega_i} = \mathcal{B}_\Omega$.

1.1.14 Corollary (Borel σ-algebra on $\mathbb{R}^n$)

$\mathcal{B}_{\mathbb{R}^n} = \bigotimes_{i=1}^{n} \mathcal{B}_{\mathbb{R}}$

Proof.

Follows from the previous proposition.

1.2 Measures

Note that we will define a measure with a σ-algebra as its domain, however very often we may want to start with a measure defined on an algebra $\mathcal{A}$ and then extend it to a measure on $\sigma(\mathcal{A})$. We define a measure only defined on an algebra to be a *pre-measure*. Similarly, one can begin with a definition of a measure on a class of subsets of Ω that only form a *semi-algebra*.

1.2.1 Measure Spaces:

Let Ω be a set equipped with a σ-algebra $\mathcal{F}$. A measure μ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that:

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

$$2)\ \{E_n\}_{n\in\mathbb{N}},\ E_j \cap E_k = \emptyset,\ j \neq k \implies \mu(\bigcup_{n\in\mathbb{N}} E_n) = \sum_{n\in\mathbb{N}} \mu(E_n)$$

We call 2) countable additivity, and refer to $(\Omega, \mathcal{F}, \mu)$ as a *measure space*.

The following is some standard terminology concerning the "size" of μ on $(\Omega, \mathcal{F})$. If $\mu(\Omega) < \infty$, we say that μ is finite. If $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for all i, we say that μ is σ-finite.

1.2.2 Lemma (Disjointification):

From any sequence $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega)$, one can construct a pairwise disjoint sequence $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega)$ defined by $B_1 = A_1$ and for $n \geq 2$:

$$B_n = A_n \cap (\bigcup_{i=1}^{n-1} A_i)^c$$

with the same union

$$\bigcup_{n\in\mathbb{N}} A_n = \dot{\bigcup}_{n\in\mathbb{N}} B_n$$

where $\dot{\bigcup}$ denotes a disjoint union.

Proof.

(Verify This!)

1.2.3 Theorem (Properties of Measures):

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We have:

$$1)\ A, B \in \mathcal{F},\ A \subset B \implies \mu(A) \leq \mu(B)$$

$$2)\ A \subset \bigcup_{i=1}^{\infty} A_i \implies \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

$$3)\ A_i \uparrow A \implies \mu(A_i) \uparrow \mu(A)$$

$$4)\ A_i \downarrow A \implies \mu(A_i) \downarrow \mu(A)$$

Proof:

1) Note $B = A \cup (B \cap A^C)$ and use finite additivity:

$$\mu(B) = \mu(A) + \mu(B \cap A^c) \geq \mu(A)$$

2) Set $A'_j = A_j \cap A$, $B_1 = A'_1$ and $B_j = A'_j \cap (\bigcup_{j=1}^{n-1} A'_j)^c$ for $j > 1$

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

which follow from our disjointification lemma and 1).

3) Set $A_0 = \emptyset$,

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i \cap A_i^c) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i \cap A_i^c) = \lim_{n\to\infty} \mu(A_n).$$

4) Set $B_j = A_1 \cap A_j^c$ and observe $\mu(A_1) = \mu(B_j) + \mu(A_j)$ then by 3):

$$\mu(A_1) = \mu(\bigcap_{i=1}^{\infty} A_i) + \lim_{i\to\infty} \mu(B_i) = \mu(\bigcap_{i=1}^{\infty} A_i) + \lim_{i\to\infty} (\mu(A_1) - \mu(A_i))$$

Subtract $\mu(A_1)$ and we get the desired result.

1.2.4 Dirac Measure

Let $(\Omega, \mathcal{F})$ be a measurable space. Let $w \in \mathcal{F}$, then the *Dirac Measure* at w is

$$\delta_w(E) := \begin{cases} 0, & w \notin E \\ 1, & w \in E \end{cases}$$

1.2.5 Discrete Measure

If $(\Omega, \mathcal{F}, \mu)$ is a measure space, then an *atom* is a subset $A \subset \Omega \in \mathcal{F}$ such that:

$$1)\ \mu(A) = A > 0.$$

$$2)\ \forall B \subset A, \mu(B) = A \text{ } or \text{ } \mu(B) = 0$$

The measure space is called *discrete* if we can write:

$$\Omega = Z \,\dot\cup\, (\dot\bigcup_{n\in\mathbb{N}} A_n)$$

where $\mu(Z) = 0$ and $\{A_n\}_{n\in\mathbb{N}}$ is a collection of atoms.

We say μ is *discrete* if and only if it is a series of dirac measures.

1.2.6 Counting Measure

The *counting measure* μ assigns to any set X the cardinality of that set X:

$$\mu(X) := |X|$$

The *counting measure* of any infinite set is ∞.

1.2.7 Lebesgue Measure (sketch and facts)

The *Lebesgue measure* on $\mathbb{R}^n$ is a model of "length", "area", "volume" for n = 1, 2, 3 respectively. Many texts tend to define the *Lebesgue Measure* of an interval in $\mathbb{R}$ as it's length. More importantly, the *Lebesgue measure* has the property of *translation invariance* meaning for any set $U \subset \mathbb{R}^n$, and any $v \in \mathbb{R}^n$:

$$\mu(U) = \mu(U + v)$$

The *Lebesgue measure* of a ball is

$$\mathcal{L}^n(B_{x,r}) = \mathcal{L}^n(B_{0,1})r^n = \frac{2\pi^{n/2}r^n}{n\Gamma(n/2)} = \frac{1}{n}\mathcal{L}^n(\mathbb{S}^{n-1})r^n$$

where

$$\mathcal{L}^n(\mathbb{S}^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$$

is the area of $\mathbb{S}^{n-1}$ which is the sphere of radius r = 1 in $\mathbb{R}^n$. The classical measure of the ball can be extended in a countably additive way to a sigma algebra containing all balls which makes integration theory relatively painless.

This discussion is continued more once we learn more about Borel Measures.

1.3 Vitali Sets

1.3.1 Why not the Power Set?

A natural question to ask is why not take the power set as our sigma-algebra each time? If our space Ω is at most countably infinite, then we have no issues in doing this.

However, if we are interested in uncountably infinite spaces like $\Omega = \mathbb{R}$ (or $\bar{\mathbb{R}}$), the prototypical counterexample is to try to define a measure that generalizes the notion of the length of an interval $I \subset \mathbb{R}$ on the entirety of $\mathcal{P}(\mathbb{R})$.

It is impossible to demand a set function λ satisfy:

$$1)\ \lambda : \mathcal{P}(\mathbb{R}) \to [0, \infty]$$

$$2)\ \lambda((a, b]) = b - a$$

$$3)\ A \subset \mathbb{R} \implies \forall x \in \mathbb{R} : \lambda(A) = \lambda(\{a + x : a \in A\})$$

$$4)\ \{A_i\}_{i\in\mathbb{N}} \subset 2^{\mathbb{R}},\ A_m \cap A_m = \emptyset,\ n \neq m \implies \lambda(\dot{\bigcup_{i=1}^{\infty}} A_i) = \sum_{i=1}^{\infty} \lambda(A_i)$$

Conditions 2), 3), and 4) are non negotiable for a function that is supposed to capture length. (Verify this!). We now show the existence of a set which one cannot assign a notion of "size".

1.3.2 The Vitali Set

A *coset of* $\mathbb{Q}$ in $\mathbb{R}$ to be any set of the form:

$$x + \mathbb{Q} = \{x + q \mid q \in \mathbb{Q}\}$$

where $x \in \mathbb{R}$. The cosets of $\mathbb{Q}$ form a partition of $\mathbb{R}$ and also happen to be dense in $\mathbb{R}$. We denote the collection of all the cosets of $\mathbb{Q}$ in $\mathbb{R}$ as $\mathbb{R}/\mathbb{Q}$. It turns out that these cosets can be used to create structures on $\mathbb{R}$ that violate our geometric intuition.

A subset $V \subset [0,1]$ is called a *Vitali* set if V contains a single point from each coset of $\mathbb{Q}$ in $\mathbb{R}$. We can "construct" a *Vitali set* using the *axiom of choice*, simply by choosing one element of (x+$\mathbb{Q}$) $\cap$ $[0,1]$ for each coset $x + \mathbb{Q} \in \mathbb{R}/\mathbb{Q}$. Of course, this "construction" is difficult to describe algorithmically, since we are making uncountably many arbitrary choices.

The *axiom of choice* says that every collection of non-empty, pair-wise disjoint sets has a *choice set.* A *choice set* for a set A is a set that contains exactly one member from each member of A.

We will take it as a fact that V is not Lebesgue measurable. The main takeaway of the Vitali set should be that there does not exist such a measure satisfying all 4 properties we want a measure to have.

1.4 Construction of Measures

1.4.1 Outer Measures

The abstract generalization of the notion of an "outer" area is the *outer measure* on a nonempty set Ω:

$$\mu^* : \mathcal{P}(\Omega) \to [0, \infty] :$$

that satisfies:

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

$$A \subset B \implies \mu^*(A) \leq \mu^*(B)$$

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

The most common way to obtain outer measures is to start with a family $\mathcal{E}$ of "elementary" sets on which a measure is defined (such as rectangles in the plane) and then to approximate arbitrary sets "from the outside" by countable unions of members of $\mathcal{E}$. The next proposition gives a precise construction.

1.4.2 Outer Measure construction

Let $\mathcal{E} \subset \mathcal{P}(\Omega)$ and f: $\mathcal{E} \to [0, \infty]$ such that $\emptyset,\ \Omega \in \mathcal{E}$ and $f(\emptyset) = 0$. For any $A \subset \Omega$ define

$$\mu^*(A) = \inf\{\sum_{n=1}^{\infty} f(E_n) : E_n \in \mathcal{E},\ A \subset \bigcup_{n=1}^{\infty} E_n\}$$

then μ^* is an outer measure.

Proof.

For any $A \in \Omega$ there exists a covering $\{E_i\}_{i=1}^{\infty} \subset \mathcal{E}$, $A \subset \bigcup_{i=1}^{\infty} E_i$. We can take $E_i = \emptyset$ for all i showing $\mu^*(\emptyset) = 0$. Note that $\mu^*(A) \leq \mu^*(B)$ for $A \subset B$ because the set in which the infimum is taken over in the definition of $\mu^*(A)$ includes the corresponding set in $\mu^*(B)$, Suppose $\{A_i\}_{i=1}^{\infty} \subset \mathcal{P}(\Omega)$ and let $\epsilon > 0$. For each i, there exists $\{E_i^k\}_{k=1}^{\infty} \subset \mathcal{E}$ such that $A_i \subset \bigcup_{k=1}^{\infty} E_i^k$ and $\sum_{k=1}^{\infty} f(E_i^k) \leq \mu^*(A_i) + (2^{-i})\epsilon$

Now if $\bigcup_{i=1}^{\infty} A_i = A$, we have $A \subset \bigcup_{i,k=1}^{\infty} E_i^k$ and $\sum_{i,k} f(E_i^k) \leq \sum_i \mu^*(A_i) + \epsilon$. Since ϵ is arbitrary we have shown countable subadditivity holds.

1.4.3 Caratheodory-measurability

If μ^* is an outer measure defined on $\mathcal{P}(\Omega)$, a set $A \subset \Omega$ is said to be μ^*-*measurable* (or *Caratheodory measurable*) if:

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset \Omega.$$

Note one can define an *"inner measure"* μ_* by $\mu_*(E) = \mu^*(\Omega) - \mu^*(E^c)$. If μ^* was induced from a countably additive measure defined on some algebra of sets in Ω, then a subset of Ω will be *Caratheodory measurable* if and only if its outer and inner measures agree.

The point to emphasize here is that the concept of "inner area" is redundant and can be defined in terms of the outer measure. Take note of this definition as it is fundamental in obtaining measures from outer measures.

1.4.4 π and λ systems

We start out with some preliminary definitions of different collections of sets. A collection $\mathcal{P}$ is said to be a π-system if:

$$A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$$

and a collection $\mathcal{L}$ is said to be a λ-system if:

$$1)\ \Omega \in \mathcal{L}$$

$$2)\ A, B \in \mathcal{L},\ A \subset B \implies B \cap A^c \in \mathcal{L}$$
$$3)\ A_n \in \mathcal{L},\ A_n \uparrow A \implies A \in \mathcal{L}$$

A class $\mathcal{C}$ of intervals in $\mathbb{R}$ is a π-system whereas the class of all open discs (balls) in $\mathbb{R}^2$ is not. By convention, we assume that these classes both contain $\emptyset$. To see this for $\mathbb{R}^2$ note that $\mathbb{B}((0,0),1) \cap \mathbb{B}((1,0),1)$ is not a ball.

An important thing to keep in mind is that every σ-algebra is a λ-system, but an algebra need not be a λ-system.

1.4.5 Theorem ($\pi - \lambda$):

If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof:

We first propose that if $\mathcal{L}(\mathcal{P})$ is the λ-system generated by $\mathcal{P}$, then $\mathcal{L}(\mathcal{P})$ is a σ-algebra. To see this note, $\sigma(\mathcal{P}) \subset \mathcal{L}(\mathcal{P}) \subset \mathcal{L}$. It suffices to show that $\mathcal{L}(\mathcal{P})$ is a π-system.

First, define $\mathcal{G}_A := \{B : A \cap B \in \mathcal{L}(\mathcal{P})\}$.

It follows that $\Omega \in \mathcal{G}_A$ since $A \in \mathcal{L}(\mathcal{P})$. Further, if:

$$B, C \in \mathcal{G}_A \text{ such that } C \subset B$$

the

$$A \cap (B \cap C^c) = (A \cap B) \cap (A \cap C)^c \in \mathcal{L}(\mathcal{P})$$

since $A \cap B$ and $A \cap C \in \mathcal{L}(\mathcal{P})$ and $\mathcal{L}(\mathcal{P})$ is a λ-system.

Lastly,

$$B_n \in \mathcal{G}_A \text{ and } B_n \uparrow B$$

then $A \cap B_n \uparrow A \cap B \in \mathcal{L}(\mathcal{P})$ since $A \cap B_n \in \mathcal{L}(\mathcal{P})$. Since $\mathcal{P}$ is a π-system,

$$A \in \mathcal{P} \implies \mathcal{P} \subset \mathcal{G}_A \text{ hence } \mathcal{L}(\mathcal{P}) \subset \mathcal{G}_A.$$

Now,

$$A \in \mathcal{L}(\mathcal{P}) \implies \mathcal{P} \subset \mathcal{G}_A \text{ so we get that } \mathcal{G}_A \subset \mathcal{L}(\mathcal{P}) \text{ as desired.}$$

Note that $\pi - \lambda$ theorem has an equivalent statement called the *monotone class theorem.*

1.4.6 Complete Measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We say that $N \subset \mathcal{F}$ is a *null set* if $\mu(N) = 0$.

If a statement $x \in \Omega$ is true except for some x in a null set we say that the statement is true *almost everywhere.* Analogously, if our measure space is a probability space, we say the statement is true *almost surely.*

We say $(\Omega, \mathcal{F}, \mu)$ is *complete* if $\mathcal{F}$ contains all subsets of sets of measure 0. Completeness can always be achieved by "en-largening" $\mathcal{F}$ as described by the following theorem below:

1.4.7 Theorem (Completion of Measure Spaces):

Suppose $(\Omega, \mathcal{M}, \mu)$ is a measure space. Let $\mathcal{N} := \{N \in \mathcal{M} : \mu(N) = 0\}$ and $\bar{\mathcal{M}} := \{E \cup F : E \in \mathcal{F},\ F \subset N \text{ for some } N \in \mathcal{N}\}$. Then $\bar{\mathcal{M}}$ is a σ-algebra, and there is a unique extension $\bar{\mu}$ of μ to a complete measure on $\bar{\mathcal{M}}$.

Proof.

First note that since $\mathcal{M}$ and $\mathcal{N}$ are closed under countable unions, $\bar{\mathcal{M}}$ is too. If $E \cup F \in \bar{\mathcal{M}}$ where $E \in \mathcal{M}$ and $F \subset N$ for some $N \in \mathcal{N}$, we can then assume that $E \cap N = \emptyset$. Otherwise we can replace F and N by $F \cap E^c$ and $N \cap E^c$ respectively. We have that $E \cup F = (E \cup N) \cap (N^c \cup F)$, so $(E \cup F)^c = (E \cup N)^c \cup (N \cap F^c)$ But $(E \cup N)^c \in \mathcal{M}$ and $N \cap F^c \subset N$ so $(E \cup F)^c \in \bar{\mathcal{M}}$ implying $\bar{\mathcal{M}}$ is a σ-algebra.

If $E \cup F \in \bar{\mathcal{M}}$ where $E \in \mathcal{M}$ and $F \subset N$ for some $N \in \mathcal{N}$, we set $\bar{\mu}(E \cup F) = \mu(E)$. We can check that this is well defined, since if $E_1 \cup F_1 = E_2 \cup F_2$, where $F_j \subset N_j \in \mathcal{N}$ then $E_1 \subset E_2 \cup N_2$ and so $\mu(E_1) \leq \mu(E_2) + \mu(N_2) = \mu(E_2)$, and likewise $\mu(E_2) \leq \mu(E_1)$. $\bar{\mu}$ is a complete measure on $\bar{\mathcal{M}}$ and is the only measure on $\bar{\mathcal{M}}$ that extends μ. (Verify This!)

We call $\bar{\mu}$ the *completion* of μ and $\bar{\mathcal{M}}$ is called the *completion* of $\mathcal{M}$ with respect to μ.

1.4.8 Theorem (Caratheodory Extension):

If μ^* is an outer measure defined on $\mathcal{P}(\Omega)$ the collection of μ^*-measurable sets is a σ-algebra and the restriction of μ^* to this collection is a complete measure.

Alternatively, if μ is a σ-finite measure defined on an algebra $\mathcal{A}$, then μ has a unique extension to $\sigma(\mathcal{A})$.

Proof.

Let $\mathcal{P}$ be a π-system. If μ_1 and μ_2 are measures (defined on $\mathcal{F}_1$ and $\mathcal{F}_2$) that agree on $\mathcal{P}$ and there exists $\{A_n\}_{n \in \mathbb{N}} \in \mathcal{P}$ with $A_n \uparrow \Omega$ and $\mu_i(A_n) < \infty$, then μ_1 and μ_2 agree on $\sigma(\mathcal{P})$. To prove uniqueness proceed as follows:

Let $A \in \mathcal{P}$ such that $\mu_1(A) = \mu_2(A) < \infty$,

$$\mathcal{L} = \{B \in \sigma(\mathcal{P}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}$$

Since $A \in \mathcal{P}$, we have $\mu_1(A) = \mu_2(A)$ and $\Omega \in \mathcal{L}$. If $C \subset B \in \mathcal{L}$:

$$\mu_1(A \cap (B \cap C^c)) = \mu_1(A \cap B) - \mu_1(A \cap C)$$

$$= \mu_2(A \cap B) - \mu_2(A \cap C) = \mu_2(A \cap (B \cap C^c))$$

If $B_n \in \mathcal{L},\ B_n \uparrow B$:

$$\mu_1(A \cap B) = \lim_{n\to\infty} \mu_1(A \cap B_n) = \lim_{n\to\infty} \mu_2(A \cap B_n) = \mu_2(A \cap B)$$

by continuity from below.

$\pi - \lambda$ implies $\sigma(\mathcal{P}) \subset \mathcal{L}$. Letting $A_n \in \mathcal{P},\ A_n \uparrow \Omega$, $\mu_1(A_n) = \mu_2(A_n) < \infty$, using the last result along with measure continuity we reach the uniqueness conclusion.

For existence, we must show that an arbitrary measure defined on an algebra $\mathcal{A}$ has an extension to $\sigma(\mathcal{A})$. Let μ^* be an outer measure. We now state two useful lemmas that are sufficient to show existence:

1) If $A \in \mathcal{A}$, then $\mu^*(A) = \mu(A)$ and A is μ^*-measurable.

2) The class $\mathcal{A}^*$ of μ^*-measurable sets is a σ-algebra and the restriction of μ^* to $\mathcal{A}^*$ is a measure.

To prove 1), without loss of generality assume $\mu^*(F) < \infty$ and let $\epsilon > 0$, there exists a $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{A}$ such that $F \subset \bigcup_{i=1}^{\infty} B_i$:

$$\sum_{i=1}^{\infty} \mu(B_i) \leq \mu^*(F) + \epsilon$$

Since μ is additive on $\mathcal{A}$, $\mu = \mu^*$ on $\mathcal{A}$ we have:

$$\mu(B_i) = \mu^*(B_i \cap A) + \mu^*(B_i \cap A^c)$$

Finally, summing over i and using the sub-additivity of μ^*:

$$\mu^*(F \cap A) + \mu^*(F^c \cap A) \leq$$

$$\sum_{i=1}^{\infty} \mu^*(B_i \cap A^c) + \sum_{i=1}^{\infty} \mu^*(B_i \cap A)$$

$$\leq \mu^*(F) + \epsilon.$$

which proves 1) since ϵ was arbitrary.

We note that if E is μ^*-measurable then E^c is too. Now if E_1 and E_2 are μ^*-measurable, then both their intersection and union are μ^*-measurable. To see this let G be any subset of Ω:

$$\mu^*(G \cap (E_1 \cup E_2)) + \mu^*(G \cap (E_1 \cup E_2)^c)$$

$$\leq \mu(G \cap E_1) + \mu(G \cap E_1^c \cap E_2) + \mu(G \cap E_1^c \cap E_2^c)$$
$$= \mu^*(G \cap E_1) + \mu^*(G \cap E^c) = \mu^*(G)$$

This proves the conclusion for the union. Observe $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$ and use the μ^*-measurability for complements to prove the conclusion for the intersection.

Let $G \subset \Omega$ and $E_1, \ldots, E_n$ be disjoint μ^*-measurable sets. Then

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

To see this, define $F_m := \bigcup_{i \le m} E_i$ where E_n is μ^*-measurable, $E_n \subset F_n$ and $F_{n-1} \cap E_n = \emptyset$. so

$$\begin{aligned}\mu^*(G \cap F_n) &= \mu^*(G \cap F_n \cap E_n) + \mu^*(G \cap F_n \cap E_n^c) \\ &= \mu^*(G \cap E_n) = \mu^*(G \cap F_{n-1})\end{aligned}$$

Induction then gives us the desired "μ^*-additivity result".

Now, we must show that if $\{E_i\}_{i \in \mathbb{N}}$ is μ^*-measurable, then $\bigcup_{i \in \mathbb{N}} E_i$ is μ^*-measurable. To this end we appeal to the disjointifcation lemma, $E_i' = E_i \cap (\bigcap_{j<i} E_j^c)$. Our previous results imply E_i' is μ^*-measurable, and let $F_n = \bigcup_{i=1}^{n} E_i$ which is also μ^* measurable. By monotonicity and μ^*-additivity, we have:

$$\begin{aligned}\sum_{i=1}^{n} \mu^*(G \cap E_i) + \mu^*(G \cap E^c) &= \mu^*(G \cap F_n) + \mu^*(G \cap E^c) \\ &\le \mu^*(G \cap F_n^c) + \mu^*(G \cap F_n) = \mu^*(G)\end{aligned}$$

Let $n \to \infty$ and subadditivity gives us:

$$\mu^*(G \cap E^c) + \mu^*(G \cap E) \le \sum_{i=1}^{\infty} \mu^*(G \cap E_i) + \mu^*(G \cap E^c) \le \mu^*(G)$$

Finally, we show countable additivity. Let $E_1, \ldots, E_n$ be disjoint and μ^*-measurable. Let $F_n = \bigcup_{i=1}^{n} E_i$. By monotonicity and μ^*-additivity,

$$\sum_{i=1}^{n} \mu^*(E_i) = \mu^*(F_n) \le \mu^*(E)$$

Letting $n \to \infty$ and subadditivity gives us our result!

Remark:

If $\mu^*(B) = 0$ then for any $E \subset \Omega$ we have

$$\mu^*(E) \le \mu^*(E \cap B) + \mu^*(E \cap B^c) \le \mu^*(E).$$

Hence μ^* restricted to the collection of μ^*-measurable sets is a complete measure.

1.4.9 Pre-Measures

One of the immediate applications of Caratheodory's Theorem is the extension of the domain of a measure from algebras to sigma-algebras. More precisely, let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be an algebra. A function $\mu_0 : \mathcal{A} \to [0, \infty]$ is called a *pre-measure* if

$$1) \ \mu_0(\emptyset) = 0$$

$$2) \ \{A_i\}_{i \in \mathbb{N}}, \ A_n \cap A_m = \emptyset; \ \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \implies \mu_0(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu_0(A_i)$$

The important thing to note is that μ_0 induces an outer measure on Ω, namely:

$$\mu^*(E) = \inf\{\sum_{n=1}^{\infty} \mu_0(A_n) : A_n \in \mathcal{A}, \ E \subset \bigcup_{n=1}^{\infty} A_n\}$$

1.4.10 Outer-Measure/Pre-measure relation

If μ_0 is a premeasure on $\mathcal{A}$ and μ^* is defined by above then the restriction of μ^* to $\mathcal{A}$ is μ_0 i.e. $\mu^*|\mathcal{A} = \mu_0$. It also follows that every set in $\mathcal{A}$ is μ^*-measurable.

Proof.

(Verify this!)

Can also be found in Folland (pg. 31).

1.4.11 Theorem (Premeasure-Extension):

Let $\mathcal{A} \subset \mathcal{P}(\Omega)$ be an algebra, μ_0 a premeasure on $\mathcal{A}$, and $\mathcal{F} = \sigma(\mathcal{A})$. There exists a measure μ defined on $\mathcal{F}$ whose restriction to $\mathcal{A}$ is μ_0 namely $\mu = \mu^*|\mathcal{F}$ where μ^* is given by the previous proposition. If ν is another measure defined on $\mathcal{F}$ then $\nu(E) \le \mu(E)$ for all $E \in \mathcal{F}$ with equality when $\mu(E) < \infty$. If μ_0 is σ-finite, then μ is the unique extension of μ_0 to a measure defined on $\mathcal{F}$.

Proof.

The first assertion follows from Caratheordory's Theorem and the previous pre-measure proposition. If $E \in \mathcal{F}$ and $E \subset \bigcup_{i=1}^{\infty} A_i$ where $A_i \in \mathcal{A}$ then $\nu(E) \le \sum_{i=1}^{\infty} \nu(A_i) = \sum_{i=1}^{\infty} \mu_0(A_i)$ hence $\nu(E) \le \mu(E)$. If we set $A = \bigcup_{i=1}^{\infty} A_i$, we have:

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

Let $\epsilon > 0$. If $\mu(E) < \infty$ we are able to choose $A_i's$ such that $\mu(A) < \mu(E) + \epsilon$ implying $\mu(A \cap E^c) < \epsilon$ and

$$\mu(E) \le \mu(A) = \nu(A) = \nu(E) + \nu(A \cap E^C) \le \nu(E) + \mu(A \cap E^C) \le \nu(E) + \epsilon$$

Since ϵ is arbitrary, $\mu(E) = \nu(E)$. Finally suppose $\Omega = \bigcup_{i=1}^{\infty} A_i$ with $\mu_0(A_i) < \infty$, with $A_i's$ assumed disjoint. Then for any $E \in \mathcal{F}$

$$\mu(E) = \sum_{i=1}^{\infty} \mu(E \cap A_i) = \sum_{i=1}^{\infty} \nu(E \cap A_i) = \nu(E)$$

showing $\nu = \mu$.

The proof of the theorem yields more than the statement. μ_0 can be extended to a measure on the algebra $\mathcal{F}^*$ of all μ^*-measurable sets.

1.4.12 Sigma-finite completion

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let μ^* be the outer measure induced by μ according to the pre-measure proposition. Define $\mathcal{F}^*$ to be the sigma algebra of all μ^*-measurable sets, and $\bar{\mu} = \mu^* | \mathcal{F}^*$ (restriction of μ^* to $\mathcal{F}^*$). It follows if μ is σ-finite then $\bar{\mu}$ turns out to be the completion of μ.

Proof.

(Verify This!)

1.5 Borel Measures

We now dissuss a large family of measures on $\mathbb{R}$ whose domain is the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$. We call such measures *Borel measures* on $\mathbb{R}$. The general guiding idea is that the measure of an interval is it's length. The generalization to $\mathbb{R}^n$ is also relatively straightforward.

Let $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mu)$ be a measure space such that μ is a finite Borel measure. Define the distribution function of μ as $F(x) = \mu((-\infty, x])$. F is increasing by monotonicity and right continuous by the continuity from below property since $(x, \infty] = \bigcap_{n=1}^{\infty}(-\infty, x_n]$ whenever $x_n \downarrow x$.

If $a < b$, $(-\infty, b] = (-\infty, a] \cup (a, b]$ we get:

$$\mu((a, b]) = F(b) - F(a)$$

Our goal is now to construct a measure μ starting from a distribution function F.

The building blocks for this theory are left-open, right closed intervals (i.e sets of the form $(a, b]$ or (a, ∞) or $\emptyset$, where $-\infty \leq a < b < \infty$. We call these sets h-interals. It follows that if $\mathcal{A}$ is the collection of finite disjoint unions of h-intervals, $\mathcal{A}$ is an algebra, and $\sigma(\mathcal{A}) = \mathcal{B}_{\mathbb{R}}$.

1.5.1 "H-interval" Pre-Measure

Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. If $(a_i, b_i]$ for $i = 1, \ldots, n$ are disjoint h-intervals, then μ_0 defines a pre-measure on $\mathcal{A}$ if:

$$\mu_0(\bigcup_{i=1}^{\infty}(a_i, b_i]) = \sum_{i=1}^{\infty}(F(b_i) - F(a_i)),\ \ \mu(\emptyset) = 0.$$

Proof.

We first check that μ_0 is well defined. If $\{(a_i, b_i]\}_{n=1}^{n}$ are disjoint* and $\bigcup_{i=1}^{n}(a_i, b_i] = (a, b]$ then (potentially after relabeling) we have a partition, $a = a_1 < b_1 = a_2 < b_2 = \ldots < b_n = b$, such that $\sum_{i=1}^{n}(F(b_i) - F(a_i)) = F(b) - F(a)$. More generally, if $\{I_i\}_{i=1}^{n}$ $\{J_j\}_{j=1}^{n}$ are finite sequences of disjoint h-intervals such that $\bigcup_{i=1}^{n} I_i = \bigcup_{j=1}^{n} J_j$:

$$\sum_i \mu_0(I_i) = \sum_{i,j} \mu_0(I_i \cap J_j) = \sum_j \mu_0(J_j)$$

which shows μ_0 is well defined and finitely additive.

It remains to show that if $\{I_i\}_{i=1}^{n}$ is a sequence of disjoint h-intervals with $\bigcup_{i=1}^{n} I_i = \mathcal{A}$, then countable additivity holds. Since $\bigcup_{i=1}^{\infty} I_i$ is a finite union of h-intervals, $\{I_i\}_{i=1}^{\infty}$ can be partitioned into finitely many subsequences such that the union of the intervals in each subsequence is a single *h-interval*. Considering each subsequence separately and using the finite additivity of μ_0, we can assume that $\bigcup_{i=1}^{\infty} I_i$ is a h-interval I = (a,b]. We have:

$$\sum_{i=1}^{n} \mu_0(I_i) = \mu_0(\bigcup_{i=1}^{n} I_i) \leq \mu_0(\bigcup_{i=1}^{n} I_i) + \mu_0(I \cap (\bigcup_{i=1}^{n} I_i)^c) = \mu_0(I)$$

Letting $n \to \infty$, it follows $\sum_{i=1}^{\infty} \mu_0(I) \leq \mu_0(I)$.

Suppose a and b are finite and let $\epsilon > 0$. F is right continuous, so there exists a δ such that F(a+δ) - F(a) $< \epsilon$ and if $I_i = (a_i, b_i]$ for each i there exists a δ_i (dependent on i) such that F($b_i + \delta_i$) - F(b_i) $< (2^{-i})\epsilon$. Note that [a+δ, b] $\subset \bigcup_{i=1}^{\infty}(a_i, b_i + \delta_i)$. Since [a+$\delta$, b] is compact, it follows that $\bigcup_{i=1}^{\infty}(a_i, b_i + \delta_i)$ reduces to a finite subcover by definition. By discarding $(a_i, b_i + \delta_i)$ that aren't in the finite subcover and relabeling the index i, we can assume that $(a_1, b_1), \ldots, (a_N, b_N)$ form a cover for $[a + \delta, b]$ with $b_i + \delta_i \in (a_{i+1,}, b_{i+1} + \delta_{i+1})$ for i=1, $\ldots, N$.

Now,

$$\begin{aligned} \mu_0(I) &< F(b) - F(a + \delta) + \epsilon \\ &\leq F(b_N + \delta_N) - F(a_1) + \epsilon \end{aligned}$$

$$= F(b_N + \delta_N) - F(A_N) + \sum_{i=1}^{N-1} [F(a_{i+1}) - F(a_i)] + \epsilon$$

$$\leq F(b_N + \delta_N) - F(A_N) + \sum_{i=1}^{N-1} [F(b_i + \delta_i) - F(a_i)] + \epsilon$$

$$< \sum_{i=1}^{N} [F(b_i) + (2^{-i})\epsilon - F(a_i)] + \epsilon$$

$$< \sum_{i=1}^{\infty} \mu_0(I_i) + 2\epsilon$$

Thus, since ϵ is arbitrary if a and b are finite we are done.

1.5.2 Theorem (Existence of Borel Measure on $\mathbb{R}$):

If F: $\mathbb{R} \to \mathbb{R}$ is a increasing, right continuous function, there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all a, b. If G is another increasing, right continuous function, we have $\mu_f = \mu_g$ if and only if $F - G$ is constant. Conversely, if μ is a Borel measure on $\mathbb{R}$ that is finite on all bounded Borel sets and we define:

$$F(x) = \begin{cases} \mu((0,x]), \text{ if } x > 0 \\ 0, \text{ if } x = 0 \\ -\mu((x,0]), \text{ if } x < 0 \end{cases}$$

then F is increasing and right continuous, and $\mu = \mu_F$.

This theorem can also be developed using half-open intervals of the form $[a,b)$ and left continuous functions F. We also note that if μ is a finite Borel measure on $\mathbb{R}$, then $\mu = \mu_F$ on $\mathcal{A}$ where $F(x) = \mu((-\infty, x])$ is called the *cumulative distribution function (cdf)* of μ. The *cdf* differs from the F specified in the theorem above by the constant $\mu((-\infty, 0])$.

Proof.

Each F induces a pre-measure on the algebra $\mathcal{A}$ (collection of finite disjoint h-intervals) by the previous proposition. It is clear that F and G induce the same pre-measure if and only if $F - G$ is constant, and $\mathbb{R} = \bigcup_{i=-\infty}^{\infty}(i, i+1]$ which implies that these pre-measures are σ-finite. The first two assertions follow from the *pre-measure extension theorem.* The monotonicity of μ implies the monotonicity of F, similarly the continuity of μ from above and below imply the right continuity of F for $x \geq 0$ and $x < 0$. Finally, $\mu = \mu_F$ on $\mathcal{A}$ implies $\mu = \mu_F$ on $\mathcal{B}_{\mathbb{R}}$ by the uniqueness in the *pre-measure extension theorem.*

1.5.3 Lebesgue-Stijeles Measure

The *pre-measure extension theorem* implies that for each increasing and right continuous function F, we obtain not only the Borel measure μ_F, but a complete measure $\bar{\mu}_F$ whose domain includes $\mathcal{B}_{\mathbb{R}}$. $\bar{\mu}_F$ is the completion of μ_F, with a domain that is always strictly larger than $\mathcal{B}_{\mathbb{R}}$. Usually, this complete measure is also dentoed by μ_F and is called the *Lebesgue-Stijeles* measure associated to F.

Fix a complete *Lebesgue-Stijeles* measure μ on $\mathbb{R}$ associated to the increasing, right continuous function F, and we denote $\mathcal{F}_\mu$, the domain of μ. It follows that for any $E \in \mathcal{F}_\mu$:

$$\mu(E) = \inf\{\sum_{i=1}^{\infty}[F(b_i) - F(a_i)] : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]\}$$

$$= \inf\{\sum_{i=1}^{\infty} \mu((a_i, b_i]) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]\}$$

1.5.4 Lemma (Open Lebesgue-Stijeles Cover):

For any $E \in \mathcal{F}_\mu$:

$$\mu(E) = \inf\{\sum_{i=1}^{\infty} \mu((a_i, b_i]) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]\}$$

$$= \inf\{\sum_{i=1}^{\infty} \mu((a_i, b_i)) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i)\}$$

Proof.

(Verify This!)

1.5.5 Theorem (Lebesgue-Stijeles Regularity):

The *Lebesgue-Stijeles measure* satifies some nice regularity properties consistent with our intuition from calculus. In particular, a Lebesgue-Stijeles measurable set can be approximated from both inside and outside by compact sets or open sets.

More precisely, if $E \in \mathcal{F}_\mu$, then:

$$\mu(E) = \inf\{\mu(U) : E \subset U,\ \text{U is open}\}$$

$$= \sup\{\mu(K) : K \subset E,\ \text{K is compact}\}$$

Proof.

By the previous lemma, for any ϵ there exist (a_i, b_i)'s such that $E \subset \bigcup_{i=1}^{\infty}(a_i, b_i)$ and $\sum_{i=1}^{\infty} \mu((a_i, b_i)) \leq \mu(E) + \epsilon$. If $U = \bigcup_{i=1}^{\infty}(a_i, b_i)$ then U is open, $E \subset U$, and $\mu(U) \leq \mu(E) + \epsilon$. Note $\mu(E) \leq \mu(U)$ whenever $E \subset U$, so we have shown the first equality.

Suppose first that E is bounded. If E is closed, then E is compact by *Heine-Borel* and the equality is obvious. Let $\epsilon > 0$ and choose an open set U, $\bar{E} \cap E^c \subset U$ such that $\mu(U) \leq \mu(\bar{E} \cap E^c) + \epsilon$. Define $K = \bar{E} \cap U^c$. K is compact, $K \subset E$ and it follows that

$$\mu(E) - \epsilon \leq \mu(E) - \mu(U) + \mu(\bar{E} \cap E^c)$$

$$\leq \mu(E) - (\mu(U) - \mu(U \cap E^c)) = \mu(E) - \mu(E \cap U)$$

$$= \mu(K)$$

If E is unbounded, define $E_i = E \cap (i, i+1]$. For any $\epsilon > 0$, there exists compact $K_i \subset E_i$ with $\mu(E_i) - (\frac{2^{-|i|}}{3})\epsilon \leq \mu(K_i)$. Define $H_n = \bigcup_{i=-n}^{n} K_i$. H_n is compact, $H_n \subset E$ and $\mu(\bigcup_{i=-n}^{n} E_i) - \epsilon \leq \mu(H_n)$. Note $\mu(E) = \lim_{n\to\infty} \mu(\bigcup_{i=-n}^{n} E_i)$ and the result follows.

1.5.6 Theorem (Borel Sets and Measure Zero):

If $E \subset \mathbb{R}$, the following are equivalent:

$$1)\ E \in \mathcal{F}_\mu$$

2) $E = V \cap N_1$ where V is a countable intersection of open sets and $\mu(N_1) = 0$

3) $H = V \cup N_2$ where H is a countable union of closed sets $\mu(N_2) = 0$

Proof.

(Verify this!)

This theorem says that all Borel measurable sets (or more generally sets in $\mathcal{F}_\mu$) are of a reasonably simple form modulo sets of measure zero.

1.5.7 Open interval approximation:

If $E \in \mathcal{F}_\mu$ and $\mu(E) < \infty$, then for every $\epsilon > 0$, there is a set A that is a finite union of disjoint open intervals such that

$$\mu((A \cap B^c) \cup (A^c \cap B)) < \epsilon$$

Proof.

(Verify this!)

1.6 Lebesgue Measure

The *Lebesgue measure* denoted m is the *completion* of the *Borel Measure*. The domain of m is called the class of *Lebesgue measurable sets* denoted as $\mathcal{L}$. We also refer to the restriction of m to $\mathcal{B}_{\mathbb{R}}$ as the *Lebesgue measure*. Define:

$$1)\ E + s := \{x + s : x \in E\}$$

$$2)\ rE := \{rx : x \in E\}$$

where $E \subset \mathbb{R}$, *and* $s, r \in \mathbb{R}$, as *translation invariance* and *"simple" behavior under dilations*.

1.6.1 Theorem (Lebesgue Translation and Dilation):

If $E \in \mathcal{L}$, then $E + s$ and $rE \in \mathcal{L}$ for all $s, r \in \mathbb{R}$. Moreover $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$

Proof.

Since the collection of open intervals is invariant under translations and dilations, $\mathcal{B}_{\mathbb{R}}$ is too. For $E \in \mathcal{B}_{\mathbb{R}}$, define $m_s(E) = m(E + s)$ and $m^r(E) = m(rE)$ then $m_s(E)$ and $m^r(E)$ agree with m and $|r|m$ on finite unions of intervals, hence on $\mathcal{B}_{\mathbb{R}}$ by our *pre-measure extension* theorem. In particular, if $E \in \mathcal{B}_{\mathbb{R}}$ and $m(E) = 0$, then $m(E + s) = m(rE) = 0$ which implies that the class of sets of *Lebesgue measure* zero is preserved by translations and dilations. It also follows that the collection of *Lebesgue measurable sets* (members of which are a union of a *Borel set* and a *Lebesgue* null set) is preserved by translations and dilations and that $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$ for all *Lebesgue* measurable sets E.

1.6.2 Measure/Topological relation of subsets of $\mathbb{R}$

Consider the following facts. Every singleton set in $\mathbb{R}$ has *Lebesgue measure* zero, and hence so does every countable subset. In particular, $m(\mathbb{Q}) = 0$. The most surprising fact is that one can find a countable dense subset of $\mathbb{R}$ that has *Lebesgue measure* 0.

To make this precise, if we let $\{r_i\}_{i\in\mathbb{N}}$ be an enumeration of $\mathbb{Q} \cap [0, 1]$, and given $\epsilon > 0$, let I_i be the open interval centered at r_i of length $(2^{-j})\epsilon$. Then $U = (0, 1) \cap \bigcup_{i=1}^{\infty} I_i$ is open and dense in $[0, 1]$, but $m(U) \leq \sum_{i=1}^{\infty}(2^{-j})\epsilon = \epsilon$. The relative complement $[0, 1] \cap U^c$ is closed and nowhere dense, but $1 - \epsilon \leq m(U^c)$. A set that is open and dense, hence topologically "large" in a sense, can be measure-theoretically small. On the contrary, a set that is nowhere dense, hence topologically "small" in a sense, can be measure-theoretically large. However, note that a nonempty open set cannot have *Lebesgue measure* zero.

The Lebesgue null sets include not only all countable sets but examples of sets having the cardinality of the continuum (cardinality of $\mathbb{R}$). The standard example is the Cantor set, which is also of interest for other reasons.

1.6.3 Cantor Set

To be continued.

2 Integration Theory

2.1 Measurable Functions

On any measure space there is an obvious notion of an integral for functions that are, in a suitable sense, locally constant with the nice property that it can be extended to an integral for more general functions. Any set function $f : X \to Y$ induces a mapping $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ called the *inverse image*. For any $E \subset Y$, we define :

$$f^{-1}(E) = \{x \in X : f(x) \in E\}$$

which preserves unions, intersections, and complements. Thus, if $\mathcal{Y}$ is a σ-algebra on Y, $\{f^{-1}(E) : E \in \mathcal{Y}\}$ is a σ-algebra on X. If $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are measurable spaces, a mapping $f : X \to Y$ is called $(\mathcal{X}, \mathcal{Y})$-*measurable* if $f^{-1}(E) \in \mathcal{X}$ for all $E \in \mathcal{Y}$. Note that the composition of measurable functions is measurable. In particular $f : \mathbb{R} \to \mathbb{C}$ is *Lebesgue measurable* if it is $(\mathcal{L}, \mathcal{B}_{\mathbb{C}})$-measurable, and *Borel measurable* if it is $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{C}})$-measurable .

Warning: if $f, g : \mathbb{R} \to \mathbb{R}$ are *Lebesgue measurable*, it does not follow that $f \circ g$ is *Lebesgue* measurable, even if g is assumed continuous! More specifically if $E \in \mathcal{B}_{\mathbb{R}}$, we have $f^{-1}(E) \in \mathcal{L}$, but there is no guarantee that $g^{-1}(f^{-1}(E)) \in \mathcal{L}$ unless $f^{-1}(E) \in \mathcal{B}_{\mathbb{R}}$.However if f is *Borel measurable*, then $f \circ g$ is *Lebesgue measurable* or *Borel measurable* whenever g is.

2.1.1 Measurability Check

If $\mathcal{Y}$ is generated by $\mathcal{E}$, then $f : X \to Y$ is $(\mathcal{X}, \mathcal{Y})$-*measurable* if and only if $f^{-1}(E) \in \mathcal{X}$ for all $E \in \mathcal{E}$.

Proof.

(Verify This!)

This is a very useful check. Pay attention to it as it comes up and up again.

2.1.2 Measurability Equivalence

If $(X, \mathcal{F})$ is a *measurable space* and $f : X \to \mathbb{R}$, then following are equivalent:

$$a)\ f \text{ is } \mathcal{F}\text{−measurable}$$

$$b)\ f^{-1}((a, \infty)) \in \mathcal{F} \text{ for all a} \in \mathbb{R}$$
$$c)\ f^{-1}([a, \infty)) \in \mathcal{F} \text{ for all a} \in \mathbb{R}$$
$$d)\ f^{-1}((-\infty, a)) \in \mathcal{F} \text{ for all a} \in \mathbb{R}$$
$$e)\ f^{-1}((-\infty, a]) \in \mathcal{F} \text{ for all a} \in \mathbb{R}$$

Proof.

We first state a preliminary result about the *Borel* σ-algebra on $\mathbb{R}$. $\mathcal{B}_{\mathbb{R}}$ is generated by each of the following:

$$\mathcal{E}_1 = \{(a, b) : a < b\}$$

$$\mathcal{E}_2 = \{[a, b] : a < b\}$$
$$\mathcal{E}_3 = \{(a, b] : a < b\} \text{ or } \mathcal{E}_4 = \{[a, b) : a < b\}$$
$$\mathcal{E}_5 = \{(a, \infty) : a \in \mathbb{R}\} \text{ or } \mathcal{E}_6 = \{(-\infty, a) : a \in \mathbb{R}\}$$
$$\mathcal{E}_7 = \{[a, , \infty) : a \in \mathbb{R}\} \text{ or } \mathcal{E}_8 = \{[-\infty, a) : a \in \mathbb{R}\}$$

All of these sets are *Borel sets*, so by (1.1.1) we get $\sigma(\mathcal{E}_j) \subset \mathcal{B}_{\mathbb{R}}$. $\mathcal{E}_3$, $\mathcal{E}_4$ are both countable intersections of open sets. $\mathcal{E}_3$ for example, note $(a.b] = \bigcap_{n=1}^{\infty}(a + n^{-1}, b + n^{-1})$. On the other hand, recall every open set in $\mathbb{R}$ is a countable union of open intervals, so again by (1.1.1), we get $\mathcal{B}_{\mathbb{R}} \subset \sigma(\mathcal{E}_1)$. $\mathcal{B}_{\mathbb{R}} \subset \sigma(\mathcal{E}_j)$ for $j \geq 2$ can now be established by showing all open intervals are contained in $\sigma(\mathcal{E}_j)$ and applying (1.1.1). The rest of the cases are straightforward (Verify This!).

This result along with our *measurability check* give us the desired result.

2.1.3 Measurability on subsets

If $(X, \mathcal{F})$ is a *measurable* space, $E \in \mathcal{F}$, we say f is measurable on E if $f^{-1}(B) \cap E \in \mathcal{F}$ for all *Borel sets* B. Equivalently, $f|E$ is $\mathcal{F}_E$-measurable where $\mathcal{F}_E = \{G \cap E : G \in \mathcal{F}\}$. Given a set X, if $\{(Y_\alpha, \mathcal{N}_\alpha)\}_{\alpha \in A}$ is a family of measurable spaces, and $f_\alpha : X \to Y_\alpha$ for each $\alpha \in A$, there is a unique smallest σ-algebra on X with respect to which the f_α's are all measurable, namely, the σ-algebra generated by the sets $f_\alpha^{-1}(E_\alpha)$ with $E_\alpha \in \mathcal{N}_\alpha,\ \alpha \in A$. It is called the σ-algebra generated by $\{f_\alpha\}_{\alpha \in A}$. In particular, if $X = \prod_{\alpha \in A} Y_\alpha$, it is the *product/cylindrical* σ-algebra generated by the coordinate maps $\pi_\alpha : X \to Y_\alpha$

2.1.4 Measurability of maps

Let $(X, \mathcal{F})$, $(\prod_{\alpha \in A} Y_\alpha, \bigotimes \mathcal{N}_\alpha)$ be a measurable space, and $\pi_\alpha : Y \to Y_\alpha$ be coordinate maps. Then $f : X \to Y$ is $(\mathcal{F}, \mathcal{N})$-measurable if and only if $f_\alpha = \pi_\alpha \circ f$ is $(\mathcal{F}, \mathcal{N}_\alpha)$-measurable for all α.

Proof.

If f is measurable, so is each f_α since the composition of measurable maps is measurable. Conversely, if each f_α is measurable, then for all $E_\alpha \in \mathcal{N}_\alpha$,$f^{-1}(\pi_\alpha^{-1}(E_\alpha))$ $= f^{-1}(E_\alpha) \in \mathcal{F}$. Thus, f is measurable by the *measurability check*.

2.1.5 Corollary (Complex-Valued functions)

A function $f : X \to \mathbb{C}$ is $\mathcal{F}$-measurable if and only if $\mathbf{Re}[f]$ and $\mathbf{Im}[f]$ are $\mathcal{F}$-measurable.

Proof.

Note $\mathbb{C}$ is separable and homeomorphic to $\mathbb{R} \times \mathbb{R}$, hence $\mathcal{B}_\mathbb{C} = \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_\mathbb{R} \otimes \mathcal{B}_\mathbb{R}$ by the *product σ-algebra reduction* proposition.

2.1.6 Extended Real numbers

The *extended real numbers* is defined to be $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. The *Borel sets* on $\bar{\mathbb{R}}$ are $\mathcal{B}_{\bar{\mathbb{R}}} = \{E \subset \bar{\mathbb{R}} : E \cap \mathbb{R}\}$. This coincides with the usual definition of $\mathcal{B}_\mathbb{R}$ if we consider $\bar{\mathbb{R}}$ as a metric space with $d(x, y) = |\arctan(x) - \arctan(y)|$.

2.1.7 Measurability of continuous functions

If X and Y are topological spaces, every continuous function $f : X \to Y$ is $(\mathcal{B}_\mathbb{X}, \mathcal{B}_\mathbb{Y})$-measurable.

Proof.

This is a corollary to the *measurability check*. Recall that f is continuous if and only if $f^{-1}(U)$ is open in X for every open $U \subset Y$.

2.1.8 "Algebra" of measurable functions

If $f, g : X \to \mathbb{C}$ are both $\mathcal{F}$-measurable, then $f + g$ and fg are $\mathcal{F}$-measurable.

Proof.

Define $F : X \to \mathbb{C} \times \mathbb{C}$, $\phi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $\psi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ by:

$$F(x) = (f(x), g(x)),\ \phi(z, w) = z + w,\ \psi(z, w) = zw$$

Note $\mathcal{B}_{\mathbb{C}\times\mathbb{C}} = \mathcal{B}_{\mathbb{C}} \otimes \mathcal{B}_{\mathbb{C}}$ by the *measurability check*, and that F is $(\mathcal{F}, \mathcal{B}_{\mathbb{C}\times\mathbb{C}})$-measurable by the previous proposition. It follows that $f + g = \phi \circ F$ and $fg = \psi \circ F$ are $\mathcal{F}$-measurable.

2.1.9 "Limits" of measurable functions

If $\{f_i\}$ is a sequence of $\bar{\mathbb{R}}$-valued functions on $(X, \mathcal{F})$ then the functions

$$g_1(x) = \sup_i f_i(x),\ g_3(x) = \limsup_{i\to\infty} f_i(x)$$

$$g_2(x) = \inf_i f_i(x),\ g_4(x) = \liminf_{i\to\infty} f_i(x)$$

are all measurable. Furthermore, if $f(x) = \lim_{i\to\infty} f_i(x)$ exists for every $x \in X$, then f is measurable.

Proof.

We have

$$g_1^{-1}((a, \infty]) = \bigcup_{i=1}^{\infty} f_i^{-1}((a, \infty]),\ g_2^{-1}([-\infty, a))$$

so g_1 and g_2 are both measurable by the *measurability equivalence*

More generally if $h_k(x) = \sup_{i>k} f_i(x)$, then h_k is measurable for each k, so $g_3 = \inf_k h_k$ is measurable and likewise for g_4. Finally, if f exists then $f = g_3 = g_4$, so f is measurable.

From these propositions, we get that if $f, g : X \to \mathbb{R}$ are *measurable*, then so are the functions $\max(f, g)$ and $\min(f, g)$, and the fact that if $\{f_i\}_{i\in\mathbb{N}}$ is a sequence of complex-valued measurable functions, and $f(x) = \lim_{i\to\infty} f_i(x)$ exists for all x then f is measurable.

2.1.10 Decompositions of functions

For a function $f : X \to \bar{\mathbb{R}}$, define the *positive* and *negative* parts of f to be

$$f^+(x) = \max(f(x), 0),\ f^-(x) = \max(-f(x), 0)$$

Then $f = f^+ - f^-$. If f is measurable, then both the *positive* and *negative* parts of f are measurable by the discussion above. Second, if $f : X \to \mathbb{C}$, we have its *polar* decomposition:

$$f = sgn(f)|f|,\ \text{where}\ sgn(f) = \begin{cases} \frac{f}{|f|} \text{ if } f \neq 0 \\ 0 \text{ if } f = 0 \end{cases}$$

2.1.11 Simple Functions

Suppose $(X, \mathcal{F})$ is a measurable space, if $E \subset X$, the *indicator function* $\mathbb{I}_E$ (which is sometimes called the *characteristic function*) is defined by

$$\mathbb{I}_E = \begin{cases} 1 \text{ if } x \in E \\ 0 \text{ if } x \notin E \end{cases}$$

It is almost immediate that $\mathbb{I}_E$ is measurable if $E \in \mathcal{F}$. A *simple function* on X is a finite linear combination, with real (complex) coefficients, of *indicator* functions of sets in $\mathcal{F}$. We do not allow *simple functions* to assume the values $-\infty$ and $+\infty$. Equivalently, $f : X \to \mathbb{C}$ is *simple* if and only if f is measurable and the range of f is a finite subset of $\mathbb{C}$. We call

$$f = \sum_{i=1}^{n} z_i \mathbb{I}_{E_i}, \text{ where } E_i = f^{-1}(\{z_i\}), \text{ and } \mathrm{range}(f) = \{z_1, \ldots, z_n\}$$

the *standard representation* of f. It exhibits f as a linear combination with distinct coefficients, of characteristic functions of disjoint sets whose union is X. Note that if f, g are *simple*, then both $f + g$ and fg are *simple*.

The following theorem is an important approximation result that shows arbitrary measurable functions can be "nicely" approximated by *simple functions*.

2.1.12 Theorem (Approximation by simple functions):

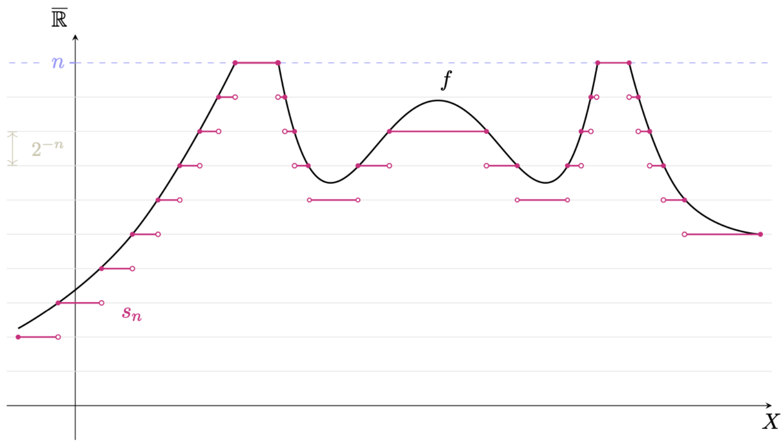
Let $(X, \mathcal{F})$ be a measurable space.

1) If $f : X \to [0, +\infty]$ is measurable, there is a sequence $\{s_n\}$ of *simple functions* such that $0 \leq s_1 \leq s_2 \leq \ldots \leq f$, $s_n \to f$ point-wise, and $s_n \to f$ uniformly on any set on which is bounded.

2) If $f : X \to \mathbb{C}$ is measurable, there is a sequence $\{s_n\}$ of *simple functions* such that $0 \leq |s_1| \leq |s_2| \leq \ldots \leq |f|$, $s_n \to f$ point-wise, and $s_n \to f$ uniformly on any set on which is bounded.

Proof.

To prove 1), note for every $n \in \mathbb{N}$, if f is unbounded above then we may cut off the height at n, and then consider partitioning $[0, n]$ into intervals of equal height 2^{-n} and then approximating f by its floor over this partition of the range. The following figure gives a pictoral demonstration for some of s_n for some $n \in \mathbb{N}$(credit: Kebeseque on Discord).



In other words for $n = 0, 1, 2, \ldots$, and $0 \leq k \leq 2^{2n} - 1$, let:

$$E_n^k = f^{-1}((k2^{-n}, (k+1)2^{-n}]) \text{ and } F_n = f^{-1}((2^{-n}, +\infty])$$

and define:

$$s_n = \sum_{k=0}^{2^n - 1} k2^{-n}(\mathbb{I}_{E_n^k}) + 2^n(\mathbb{I}_{F_n})$$

Note: $s_n \leq s_{n+1}$ for all n and $0 \leq f - s_n \leq 2^{-n}$ on the set where $f \leq 2^n$.

To prove 2), note if $f = g + ih$, we can apply part 1) to the *positive* and *negative* parts of g and h, obtaining sequences $\psi_n^+,\ \psi_n^-,\ \zeta_n^+, \zeta_n^-$ of nonnegative simple functions that increase towards $g^+,\ g^-,\ h^+,\ h^-$. Let $\phi_n = \psi_n^+ - \psi_n^- + i(\zeta_n^+ - \zeta_n^-)$ and note:

$$\psi_n^+ - \psi_n^- + i(\zeta_n^+ - \zeta^-) \to g^+ - g^- + i(h^+ - h^-) = g + ih = f$$

(This will be fleshed out later, also need to finish showing $s_n \leq s_{n+1}$ for the last part of a).)

2.1.13 Complete measurability

Suppose μ is a complete measure, if f is measurable and $f = g$ μ-almost everywhere, then g is measurable. Furthermore, if f_n is measurable for $n \in \mathbb{N}$ and $f_n \to f$ μ-almost everywhere, then f is measurable.

Proof.

Suppose μ is a complete measure, f is measurable and $f = g$ μ-almost everywhere. Define $E := \{x : f(x) \neq g(x)\}$. Suppose A is measurable and note $g^{-1}(A) = (g^{-1}(A) \cap E) \cup (g^{-1}(A) \cap E^c)$. $(g^{-1}(A) \cap E)$ is measurable as it is contained in E which has a measure of 0. Finally, note that $(g^{-1}(A) \cap E^c) = (f^{-1}(A) \cap E^c)$ by assumption hence $g^{-1}(A)$ is measurable.

The second part is a similar idea (Verify this!)

2.2 Integration of Non-Negative Functions

Fix a measure space $(\Omega, \mathcal{F}, \mu)$, and let L^+ denote the space of all measurable functions from Ω to $[0, \infty]$ If ϕ is a simple function in L^+ with standard representation $\phi = \sum_{i=1}^n a_i \mathbb{I}_{E_i}$ we define the *integral* of ϕ with respect to μ by

$$\int \phi d\mu = \sum_{i=1}^n a_i \mu(E_i)$$

2.2.1 Integral properties

Let $\phi,\ \psi$ be simple functions in L^+, then the following hold:

$$1) \text{ If } c \geq 0, \text{ then } \int c\phi = c \int \phi$$

$$2) \int (\phi + \psi) = \int \phi + \int \psi$$

$$3) \text{ If } \phi \leq \psi, \text{ then } \int \phi \leq \int \psi$$

$$4) \text{ the map } A \mapsto \int_A \phi d\mu \text{ is a measure on } \mathcal{F}.$$

Proof.

1) follows from the linearity of a sum. For 2), define $\sum_{j=1}^n a_j \mathbb{I}_{E_j}$ and $\sum_{i=1}^n b_i \mathbb{I}_{F_i}$ to be the standard representations of ϕ and ψ. Then $E_j = \bigcup_{i=1}^m (E_j \cap F_i)$ and $F_i = \bigcup_{j=1}^m (E_j \cap F_i)$. Note that $\bigcup_{j=1} E_j = \bigcup_{i=1} F_i = \Omega$ where the unions are disjoint. Hence by finite additivity,

$$\int \phi + \int \psi = \sum_{j,k} (a_j + b_k) \mu(E_j \cap F_k)$$

$$= \int \psi + \phi$$

Moreover, if $\psi \leq \phi$, then $a_j \leq b_k$ whenever $E_j \cap F_k \neq \emptyset$, so

$$\int \phi = \sum_{j,k} a_j \mu(E_j \cap F_k) \leq \sum_{j,k} b_k \mu(E_j \cap F_k) = \int \psi$$

showing 3). Finally, let $\{A_k\}$ be a disjoint sequence in $\mathcal{F}$ and $A = \bigcup_{k=1}^{\infty} A_k$:

$$\int_A \phi = \sum_j a_j \mu(A \cap E_j) = \sum_{j,k} a_j \mu(A_k \cap E_j) = \sum_k \int_{A_k} \phi$$

which establishes 4).

From this, we can extend the integral to all functions $f \in L^+$ by defining

$$\int f d\mu = \sup\{\int \phi d\mu : 0 \leq \phi \leq f, \phi \text{ simple}\}$$

2.2.2 Monotone Convergence Theorem

3 Probability Theory

3.1 Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ is a random variable, if the event $X^{-1}((-\infty, a]) := \{\omega : X(\omega) \leq a\} \in \mathcal{F}$ for each $a \in \mathbb{R}$. This is equivalent to the stronger condition: $f^{-1}(A) \in \mathcal{F}$ for all *Borel sets* $A \in \mathcal{B}_{\mathbb{R}}$

Some argue that probability theory begins and measure theory ends with the definition of independence. If one only knows undergraduate probability theory, one's intuition might come from the easily envisioned property that the occurrence or non-occurrence of an event has no effect on our estimate of the probability that an independent event will or not occur. Despite this intuitive appeal, it is important to recognize that independence is a technical concept with a technical definition which must be checked with respect to a specific probability model. There are many counterexamples to our naive "intuition". We now try to characterize independence through a few definitions.

3.2 Independence

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a fixed probability space. We say the events $A, B \in \mathcal{F}$ are *independent* if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Two real random variables X and Y are *independent* if for all $C, D \in \mathcal{B}_{\mathbb{R}}$:

$$\mathbb{P}(X \in C, Y \in D) = \mathbb{P}(X \in C)\mathbb{P}(Y \in D)$$

or in other words, the two events $A = \{A \in C\}$ $B = \{Y \in D\}$ are *independent*.

Two σ-algebra's $\mathcal{F}$ and $\mathcal{G}$ are *independent* if for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$ the events A and B are *independent*.

Let $\mathcal{C}_i \subset \mathcal{B}_{\mathbb{R}}$ for $i = 1, \ldots, n$. The classes $\mathcal{C}_i$ are *independent*, if for any choice $A_1, \ldots, A_n$, with $A_i \in \mathcal{C}_i$, we have that the events $A_1, \ldots, A_n$ are *independent*.

3.2.1 Theorem (Independence Criterion):

If $\mathcal{C}_i$ is a non empty class of events for each $i = 1, \ldots, n$, such that $\mathcal{C}_i$ is a π-system, and if $\mathcal{C}_i$'s are *independent* for $i = 1, \ldots, n$

3.3 Radon-Nikodym Theorem

Let $(\Omega, \mathcal{F}, \nu)$ be a measure space and f be a non-negative Borel measurable function. The set function:

$$\lambda(A) = \int_A f d\nu, \ A \in \mathcal{F}$$

is a measure on $(\Omega, \mathcal{F})$. Note that $\nu(A) = 0$ implies $\lambda(A) = 0$. If $\nu(A) = 0$ implies $\lambda(A) = 0$ holds for λ and ν defined on the same measurable space we say λ is *absolutely continuous* with respect to ν.

Let ν, λ be two measures on $(\Omega, \mathcal{F})$ and ν be σ-finite. If λ is absolutely continuous with respect to ν, then there exists a non-negative Borel measurable function f on Ω such that:

$$\lambda(A) = \int_A f d\nu, \ A \in \mathcal{F}$$

is a measure on $(\Omega, \mathcal{F})$. Furthermore, f is unique *almost everywhere* ν or in other words $\lambda(A) = \int_A g d\nu$ for any $A \in \mathcal{F}$, then f=g *almost everywhere* ν. We call f the *Radon-Nikodym* derivative (or *density*) of λ with respect to ν denoted as $\frac{d\lambda}{d\nu}$.

Proof.

Check Billingsley (might add conditional expectation proof)

A useful consequence of Radon-Nikodym is that if f is Borel measurable on $(\Omega, \mathcal{F})$ and $\int_A f d\nu = 0$ for any $A \in \mathcal{F}$, then $f = 0$ almost everywhere. If $\int f d\nu = 1$ for a $f \geq 0$ almost everywhere ν, then λ is given by above is a probability measure and we refer to f as it's *probability distribution function (pdf)*.

A continuous cumulative distribution function may not have a probability distribution function with respect to the Lebesgue measure. A necessary and sufficient condition is that F is *absolutely continuous* in the sense that for any $\epsilon > 0$, there exists a $\delta > 0$ such that for any finite collection of disjoint open intervals (a_i, b_i), $\sum_i (b_i - a_i) < \delta$ implies $\sum [F(b_i) - F(a_i)] < \epsilon$. *Absolutely continuity* in this sense is weaker than differentiability, but stronger than continuity.

Any discontinuous cumulative distribution function is not *absolutely continuous*, however every cumulative distribution function is differentiable almost everywhere Lebesgue measure. Thus, the important take away is that if f is the probability distribution function of F with respect to the Lebesgue measure, then f is the usual derivative of F almost everywhere Lebesgue measure and

$$F(x) = \int_{-\infty}^{x} f(y)dy, \ x \in \mathbb{R}$$

holds.

3.4 Conditional Expectation

In elementary probability, we are able to define the conditional probability of an event A given event B provided the probability of event B occurring is strictly greater than 0. However, we sometimes need a notion of "conditional" probability for events with probability 0. For example, define the event $B := \{Y = c\}$ where $c \in \mathbb{R}$ and Y is a random variable with a continuous cdf.

Let X be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{D}$ be a sub-σ-algebra of $\mathcal{F}$. The *conditional expectation* of X given $\mathcal{D}$ denoted $\mathbb{E}(X|\mathcal{D})$ is the *almost surely*-unique random variable satisfying the two conditions:

$$1)\ \mathbb{E}(X|\mathcal{D}) \text{ is measurable from } (\Omega, \mathcal{D}) \text{ to } (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

$$2) \int_D \mathbb{E}(X|\mathcal{D})d\mathbb{P} = \int_D X d\mathbb{P} \text{ for any D} \in \mathcal{D}$$

Let $B \in \mathcal{F}$. The *conditional probability* of B given $\mathcal{D}$ is defined to be:

$$\mathbb{P}(B|\mathcal{D}) = \mathbb{E}(\mathbb{I}_{\mathbb{B}}|\mathcal{D})$$

Suppose Y is measurable from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Lambda, \mathcal{G})$. The *conditional expectation of X given Y* is defined to be

$$\mathbb{E}(X|Y) = E[X|\sigma(Y)]$$

3.5 Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A (discrete-time) *filtration* is an increasing sequence of sub-σ-algebras:

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

Suppose $\{X_n\}_{n\geq 1}$ be a sequence of random variables and set

$$\mathcal{F}_n = \sigma\big(X_1, \ldots, X_n\big), \quad n \geq 1, \ \mathcal{F}_0 = \{\emptyset, \Omega\}$$

We call a sequence of integrable random variables $\{M_n\}_{n\geq 0}$ a *martingale* with respect to the filtration $\{\mathcal{F}_n\}$ if:

$$1)\ M_n \text{ is } \mathcal{F}_n\text{-measurable for each n} \geq 0$$

$$2)\ \mathbb{E}\big[\,|M_n|\,\big] < \infty \text{ for all n} \geq 0,$$

$$3)\ \mathbb{E}\big[M_n \mid \mathcal{F}_{n-1}\big] = M_{n-1} \text{ almost surely for each n} \geq 1$$

We can think of the X_i's as telling us the ith outcome of some gambling process, and M_n as the fortune of a gambler who places fair bets in varying amounts on the results of the coin tosses. 3) tells us that the expected value of the gambler's fortune at time n given all the information in the first $n-1$ flips of the coin is simply M_{n-1}, the actual value of the gambler's fortune before the nth round of the gambling process.

3.5.1 Example (Partial Sum Process):

If X_n are independent random variables with $\mathbb{E}(X_n) = 0$ for all $n \geq 1$, then the *partial sum process* given by taking $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$ for all $n \geq 1$ is a martingale with respect to $\{X_n\}_{n\geq 1}$

3.5.2 Example (Condition/Uncondition):

If X_n are independent random variables with $\mathbb{E}(X_n) = 0$, $Var(X_n) = \sigma^2$ for all $n \geq 1$, and setting $M_0 = 0$ and $M_n = S_n^2 - n\sigma^2$ for all $n \geq 1$ gives us a martingale with respect to $\{X_n\}_{n\geq 1}$

Measurability is trivial, so we focus on seperating the conditioned and unconditioned parts of the process:

$$\mathbb{E}(M_n|X_1, \ldots, X_n) = \mathbb{E}(S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2|X_1, \ldots, X_n)$$

S_{n-1}^2 is a function of $X_1, \ldots, X_{n-1}$, so its conditional expectation given $X_1, \ldots, X_{n-1}$ is just S_{n-1}^2. We note that $\mathbb{E}(X_n|X_1, \ldots, X_{n-1}) = \mathbb{E}(X_n) = 0$ which implies $\mathbb{E}(X_n^2|X_1, \ldots, X_{n-1}) = \sigma^2$.

Hence,

$$\begin{aligned}
&\mathbb{E}(S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - n\sigma^2|X_1, \ldots, X_n) \\
&= S_{n-1}^2 + 2S_{n-1}\mathbb{E}(X_n|X_1, \ldots, X_{n-1}) + \sigma^2 - n\sigma^2 \\
&= S_{n-1}^2 + 2S_{n-1}(0) + (n-1)\sigma^2 \\
&= S_{n-1}^2 + (n-1)\sigma^2
\end{aligned}$$

4 References

Athreya, K.B. and Lahiri, S.N. (2006) Measure Theory and Probability Theory. Springer, Berlin.

Durrett, R. (2019). Probability: Theory and Examples (5th ed.). Cambridge University Press.

Folland, G. B. (1999). Real Analysis: Modern Techniques and Their Applications (2nd ed.). Wiley.