Bruce K. Driver

Math 280 (Probability Theory) Lecture Notes

October 28, 2009 *File:prob.tex*

Contents

Part

Homework Problems

Math 280A Homework Problems Fall 2009

Problems are from Resnick, S. A Probability Path, Birkhauser, 1999 or from the lecture notes. The problems from the lecture notes are hyperlinked to their location.

-3.1 Homework 1. Due Wednesday, September 30, 2009

- Read over Chapter 1.
- Hand in Exercises 1.1, 1.2, and 1.3.

-3.2 Homework 2. Due Wednesday, October 7, 2009

- Look at Resnick, p. 20-27: 9, 12, 17, 19, 27, 30, 36, and Exercise 3.9 from the lecture notes.
- Hand in Resnick, p. 20-27: 5, 18, 23, 40*, 41, and Exercise 4.1 from the lecture notes.

*Notes on Resnick's #40: (i) $\mathcal{B}((0,1])$ should be $\mathcal{B}([0,1))$ in the statement of this problem, (ii) k is an integer, (iii) $r \geq 2$.

-3.3 Homework 3. Due Wednesday, October 21, 2009

- Look at Lecture note Exercises; 4.7, 4.8, 4.9
- Hand in Resnick, p. 63–70; 7* and 13.
- Hand in Lecture note Exercises: 4.3, 4.4, 4.5, 4.6, 4.10 – 4.15.

***Hint**: For #7 you might label the coupons as $\{1, 2, \ldots, N\}$ and let A_i be the event that the collector does **not** have the i^{th} – coupon after buying n - boxes of cereal.

-3.4 Homework 4. Due Wednesday, October 28, 2009

- Look at Lecture note Exercises; 5.5, 5.10.
- Look at Resnick, p. 63–70; 5, 14, 16, 19
- Hand in Resnick, p. 63–70; 3, 6, 11
- Hand in Lecture note Exercises: 5.6 – 5.9.

-3.5 Homework 5. Due Wednesday, November 4, 2009

- Look at Resnick, p. 85–90: 3, 7, 8, 12, 17, 21
- **Hand in** from Resnick, p. 85–90: 4, 6*, 9, 15, 18**.
  *Note: In #6, the random variable X is understood to take values in the extended real numbers.
  ** I would write the left side in terms of an expectation.
- Look at Lecture note Exercise 6.3, 6.7.
- **Hand in** Lecture note Exercises: 6.4, 6.6, 6.10.

-3.6 Homework 6 (Tentative). Due Wednesday, November 18, 2009

- Look at Lecture note Exercise 7.3, 7.7, 7.10, 7.15, and 7.16.
- **Hand in** Lecture note Exercises: 7.4, 7.5, 7.6, 7.9, 7.11, 7.12, 7.14
- Look at from Resnik, p. 155–166: 6, 13, 26, 37
- **Hand in** from Resnick,p. 155–166: 7, 38

Part I

Background Material

Limsups, Liminfs and Extended Limits

Notation 1.1 *The* ***extended real numbers*** *is the set* $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ *, i.e. it is* $\mathbb{R}$ *with two new points called* ∞ *and* $-\infty$*. We use the following conventions,* $\pm\infty \cdot 0 = 0$*,* $\pm\infty \cdot a = \pm\infty$ *if* $a \in \mathbb{R}$ *with* $a > 0$*,* $\pm\infty \cdot a = \mp\infty$ *if* $a \in \mathbb{R}$ *with* $a < 0$*,* $\pm\infty + a = \pm\infty$ *for any* $a \in \mathbb{R}$*,* $\infty + \infty = \infty$ *and* $-\infty - \infty = -\infty$ *while* $\infty - \infty$ *is not defined. A sequence* $a_n \in \bar{\mathbb{R}}$ *is said to converge to* ∞ $(-\infty)$ *if for all* $M \in \mathbb{R}$ *there exists* $m \in \mathbb{N}$ *such that* $a_n \geq M$ $(a_n \leq M)$ *for all* $n \geq m$*.*

Lemma 1.2. *Suppose* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *are convergent sequences in* $\bar{\mathbb{R}}$*, then:*

1. *If* $a_n \leq b_n$ *for*[1] *a.a.* n*, then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$*.*
2. *If* $c \in \mathbb{R}$*, then* $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$*.*
3. $\{a_n + b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \tag{1.1}$$
provided the right side is not of the form $\infty - \infty$*.*
4. $\{a_n b_n\}_{n=1}^\infty$ *is convergent and*
$$\lim_{n\to\infty} (a_n b_n) = \lim_{n\to\infty} a_n \cdot \lim_{n\to\infty} b_n \tag{1.2}$$
provided the right hand side is not of the for $\pm\infty \cdot 0$ *of* $0 \cdot (\pm\infty)$ *.*

Before going to the proof consider the simple example where $a_n = n$ and $b_n = -\alpha n$ with $\alpha > 0$. Then

$$\lim (a_n + b_n) = \begin{cases} \infty & \text{if } \alpha < 1 \\ 0 & \text{if } \alpha = 1 \\ -\infty & \text{if } \alpha > 1 \end{cases}$$

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{“} = \text{”} \infty - \infty.$$

This shows that the requirement that the right side of Eq. (1.1) is not of form $\infty - \infty$ is necessary in Lemma 1.2. Similarly by considering the examples $a_n = n$ and $b_n = n^{-\alpha}$ with $\alpha > 0$ shows the necessity for assuming right hand side of Eq. (1.2) is not of the form $\infty \cdot 0$.

[1] Here we use “a.a. n” as an abbreviation for almost all n. So $a_n \leq b_n$ a.a. n iff there exists $N < \infty$ such that $a_n \leq b_n$ for all $n \geq N$.

Proof. The proofs of items 1. and 2. are left to the reader.

Proof of Eq. (1.1). Let $a := \lim_{n\to\infty} a_n$ and $b = \lim_{n\to\infty} b_n$. Case 1., suppose $b = \infty$ in which case we must assume $a > -\infty$. In this case, for every $M > 0$, there exists N such that $b_n \geq M$ and $a_n \geq a - 1$ for all $n \geq N$ and this implies

$$a_n + b_n \geq M + a - 1 \text{ for all } n \geq N.$$

Since M is arbitrary it follows that $a_n + b_n \to \infty$ as $n \to \infty$. The cases where $b = -\infty$ or $a = \pm\infty$ are handled similarly. Case 2. If $a, b \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$|a - a_n| \leq \varepsilon \text{ and } |b - b_n| \leq \varepsilon \text{ for all } n \geq N.$$

Therefore,

$$|a + b - (a_n + b_n)| = |a - a_n + b - b_n| \leq |a - a_n| + |b - b_n| \leq 2\varepsilon$$

for all $n \geq N$. Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

Proof of Eq. (1.2). It will be left to the reader to prove the case where $\lim a_n$ and $\lim b_n$ exist in $\mathbb{R}$. I will only consider the case where $a = \lim_{n\to\infty} a_n \neq 0$ and $\lim_{n\to\infty} b_n = \infty$ here. Let us also suppose that $a > 0$ (the case $a < 0$ is handled similarly) and let $\alpha := \min\left(\frac{a}{2}, 1\right)$. Given any $M < \infty$, there exists $N \in \mathbb{N}$ such that $a_n \geq \alpha$ and $b_n \geq M$ for all $n \geq N$ and for this choice of N, $a_n b_n \geq M\alpha$ for all $n \geq N$. Since $\alpha > 0$ is fixed and M is arbitrary it follows that $\lim_{n\to\infty} (a_n b_n) = \infty$ as desired. ■

For any subset $\Lambda \subset \bar{\mathbb{R}}$, let $\sup \Lambda$ and $\inf \Lambda$ denote the least upper bound and greatest lower bound of Λ respectively. The convention being that $\sup \Lambda = \infty$ if $\infty \in \Lambda$ or Λ is not bounded from above and $\inf \Lambda = -\infty$ if $-\infty \in \Lambda$ or Λ is not bounded from below. We will also use the **conventions** that $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$.

Notation 1.3 *Suppose that* $\{x_n\}_{n=1}^\infty \subset \bar{\mathbb{R}}$ *is a sequence of numbers. Then*

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \inf\{x_k : k \geq n\} \text{ and} \tag{1.3}$$
$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{1.4}$$

We will also write $\underline{\lim}$ *for* $\liminf_{n\to\infty}$ *and* $\overline{\lim}$ *for* $\limsup\limits_{n\to\infty}$.

Remark 1.4. Notice that if $a_k := \inf\{x_k : k \geq n\}$ and $b_k := \sup\{x_k : k \geq n\}$, then $\{a_k\}$ is an increasing sequence while $\{b_k\}$ is a decreasing sequence. Therefore the limits in Eq. (1.3) and Eq. (1.4) always exist in $\bar{\mathbb{R}}$ and

$$\liminf_{n\to\infty} x_n = \sup_n \inf\{x_k : k \geq n\} \text{ and}$$
$$\limsup_{n\to\infty} x_n = \inf_n \sup\{x_k : k \geq n\}.$$

The following proposition contains some basic properties of liminfs and limsups.

Proposition 1.5. *Let* $\{a_n\}_{n=1}^\infty$ *and* $\{b_n\}_{n=1}^\infty$ *be two sequences of real numbers. Then*

1. $\liminf_{n\to\infty} a_n \leq \limsup\limits_{n\to\infty} a_n$ *and* $\lim_{n\to\infty} a_n$ *exists in* $\bar{\mathbb{R}}$ *iff*
$$\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n \in \bar{\mathbb{R}}.$$
2. *There is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \limsup\limits_{n\to\infty} a_n$. *Similarly, there is a subsequence* $\{a_{n_k}\}_{k=1}^\infty$ *of* $\{a_n\}_{n=1}^\infty$ *such that* $\lim_{k\to\infty} a_{n_k} = \liminf_{n\to\infty} a_n$.
3. $$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{1.5}$$
whenever the right side of this equation is not of the form $\infty - \infty$.
4. *If* $a_n \geq 0$ *and* $b_n \geq 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty}(a_n b_n) \leq \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{1.6}$$
provided the right hand side of (1.6) is not of the form $0 \cdot \infty$ *or* $\infty \cdot 0$.

Proof. 1. Since

$$\inf\{a_k : k \geq n\} \leq \sup\{a_k : k \geq n\}\ \forall\, n,$$

$$\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

Now suppose that $\liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n = a \in \mathbb{R}$. Then for all $\varepsilon > 0$, there is an integer N such that

$$a - \varepsilon \leq \inf\{a_k : k \geq N\} \leq \sup\{a_k : k \geq N\} \leq a + \varepsilon,$$

i.e.

$$a - \varepsilon \leq a_k \leq a + \varepsilon \text{ for all } k \geq N.$$

Hence by the definition of the limit, $\lim_{k\to\infty} a_k = a$. If $\liminf_{n\to\infty} a_n = \infty$, then we know for all $M \in (0, \infty)$ there is an integer N such that

$$M \leq \inf\{a_k : k \geq N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup\limits_{n\to\infty} a_n = -\infty$ is handled similarly.

Conversely, suppose that $\lim_{n\to\infty} a_n = A \in \bar{\mathbb{R}}$ exists. If $A \in \mathbb{R}$, then for every $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$ such that $|A - a_n| \leq \varepsilon$ for all $n \geq N(\varepsilon)$, i.e.

$$A - \varepsilon \leq a_n \leq A + \varepsilon \text{ for all } n \geq N(\varepsilon).$$

From this we learn that

$$A - \varepsilon \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$A \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n \leq A,$$

i.e. that $A = \liminf_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$. If $A = \infty$, then for all $M > 0$ there exists $N = N(M)$ such that $a_n \geq M$ for all $n \geq N$. This show that $\liminf_{n\to\infty} a_n \geq M$ and since M is arbitrary it follows that

$$\infty \leq \liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n.$$

The proof for the case $A = -\infty$ is analogous to the $A = \infty$ case.

2. – 4. The remaining items are left as an exercise to the reader. It may be useful to keep the following simple example in mind. Let $a_n = (-1)^n$ and $b_n = -a_n = (-1)^{n+1}$. Then $a_n + b_n = 0$ so that

$$0 = \lim_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty}(a_n + b_n) = \limsup_{n\to\infty}(a_n + b_n)$$

while

$$\liminf_{n\to\infty} a_n = \liminf_{n\to\infty} b_n = -1 \text{ and}$$
$$\limsup_{n\to\infty} a_n = \limsup_{n\to\infty} b_n = 1.$$

Thus in this case we have

$$\limsup_{n\to\infty}(a_n+b_n) < \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \text{ and}$$
$$\liminf_{n\to\infty}(a_n+b_n) > \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n.$$

■

We will refer to the following basic proposition as the monotone convergence theorem for sums (MCT for short).

Proposition 1.6 (MCT for sums). *Suppose that for each* $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^\infty$ *is a sequence in* $[0,\infty]$ *such that* $\uparrow \lim_{n\to\infty} f_n(i) = f(i)$ *by which we mean* $f_n(i) \uparrow f(i)$ *as* $n\to\infty$. *Then*

$$\lim_{n\to\infty}\sum_{i=1}^\infty f_n(i) = \sum_{i=1}^\infty f(i), \text{ i.e.}$$
$$\lim_{n\to\infty}\sum_{i=1}^\infty f_n(i) = \sum_{i=1}^\infty \lim_{n\to\infty} f_n(i).$$

We allow for the possibility that these expression may equal to $+\infty$.

Proof. Let $M :=\uparrow \lim_{n\to\infty}\sum_{i=1}^\infty f_n(i)$. As $f_n(i) \le f(i)$ for all n it follows that $\sum_{i=1}^\infty f_n(i) \le \sum_{i=1}^\infty f(i)$ for all n and therefore passing to the limit shows $M \le \sum_{i=1}^\infty f(i)$. If $N \in \mathbb{N}$ we have,

$$\sum_{i=1}^N f(i) = \sum_{i=1}^N \lim_{n\to\infty} f_n(i) = \lim_{n\to\infty}\sum_{i=1}^N f_n(i) \le \lim_{n\to\infty}\sum_{i=1}^\infty f_n(i) = M.$$

Letting $N\uparrow\infty$ in this equation then shows $\sum_{i=1}^\infty f(i) \le M$ which completes the proof. ■

Proposition 1.7 (Tonelli's theorem for sums). *If* $\{a_{kn}\}_{k,n=1}^\infty \subset [0,\infty]$, *then*

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn} = \sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Here we allow for one and hence both sides to be infinite.

Proof. First Proof. Let $S_N(k) := \sum_{n=1}^N a_{kn}$, then by the MCT (Proposition 1.6),

$$\lim_{N\to\infty}\sum_{k=1}^\infty S_N(k) = \sum_{k=1}^\infty \lim_{N\to\infty} S_N(k) = \sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}.$$

On the other hand,

$$\sum_{k=1}^\infty S_N(k) = \sum_{k=1}^\infty\sum_{n=1}^N a_{kn} = \sum_{n=1}^N\sum_{k=1}^\infty a_{kn}$$

so that

$$\lim_{N\to\infty}\sum_{k=1}^\infty S_N(k) = \lim_{N\to\infty}\sum_{n=1}^N\sum_{k=1}^\infty a_{kn} = \sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Second Proof. Let

$$M := \sup\left\{\sum_{k=1}^K\sum_{n=1}^N a_{kn} : K,N\in\mathbb{N}\right\} = \sup\left\{\sum_{n=1}^N\sum_{k=1}^K a_{kn} : K,N\in\mathbb{N}\right\}$$

and

$$L := \sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn}.$$

Since

$$L = \sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn} = \lim_{K\to\infty}\sum_{k=1}^K\sum_{n=1}^\infty a_{kn} = \lim_{K\to\infty}\lim_{N\to\infty}\sum_{k=1}^K\sum_{n=1}^N a_{kn}$$

and $\sum_{k=1}^K\sum_{n=1}^N a_{kn} \le M$ for all K and N, it follows that $L \le M$. Conversely,

$$\sum_{k=1}^K\sum_{n=1}^N a_{kn} \le \sum_{k=1}^K\sum_{n=1}^\infty a_{kn} \le \sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn} = L$$

and therefore taking the supremum of the left side of this inequality over K and N shows that $M \le L$. Thus we have shown

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn} = M.$$

By symmetry (or by a similar argument), we also have that $\sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn} = M$ and hence the proof is complete. ■

You are asked to prove the next three results in the exercises.

Proposition 1.8 (Fubini for sums). *Suppose* $\{a_{kn}\}_{k,n=1}^\infty \subset \mathbb{R}$ *such that*

$$\sum_{k=1}^\infty\sum_{n=1}^\infty |a_{kn}| = \sum_{n=1}^\infty\sum_{k=1}^\infty |a_{kn}| < \infty.$$

Then

$$\sum_{k=1}^\infty\sum_{n=1}^\infty a_{kn} = \sum_{n=1}^\infty\sum_{k=1}^\infty a_{kn}.$$

Example 1.9 (Counter example). Let $\{S_{mn}\}_{m,n=1}^{\infty}$ be any sequence of complex numbers such that $\lim_{m\to\infty} S_{mn} = 1$ for all n and $\lim_{n\to\infty} S_{mn} = 0$ for all n. For example, take $S_{mn} = 1_{m\geq n} + \frac{1}{n} 1_{m<n}$. Then define $\{a_{ij}\}_{i,j=1}^{\infty}$ so that

$$S_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}.$$

Then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \lim_{m\to\infty} \lim_{n\to\infty} S_{mn} = 0 \neq 1 = \lim_{n\to\infty} \lim_{m\to\infty} S_{mn} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

To find a_{ij}, set $S_{mn} = 0$ if $m = 0$ or $n = 0$, then

$$S_{mn} - S_{m-1,n} = \sum_{j=1}^{n} a_{mj}$$

and

$$\begin{aligned} a_{mn} &= S_{mn} - S_{m-1,n} - (S_{m,n-1} - S_{m-1,n-1}) \\ &= S_{mn} - S_{m-1,n} - S_{m,n-1} + S_{m-1,n-1}. \end{aligned}$$

Proposition 1.10 (Fatou's Lemma for sums). *Suppose that for each $n \in \mathbb{N}$, $\{h_n(i)\}_{i=1}^{\infty}$ is any sequence in $[0,\infty]$, then*

$$\sum_{i=1}^{\infty} \liminf_{n\to\infty} h_n(i) \leq \liminf_{n\to\infty} \sum_{i=1}^{\infty} h_n(i).$$

The next proposition is referred to as the dominated convergence theorem (DCT for short) for sums.

Proposition 1.11 (DCT for sums). *Suppose that for each $n \in \mathbb{N}$, $\{f_n(i)\}_{i=1}^{\infty} \subset \mathbb{R}$ is a sequence and $\{g_n(i)\}_{i=1}^{\infty}$ is a sequence in $[0,\infty)$ such that;*

1. $\sum_{i=1}^{\infty} g_n(i) < \infty$ *for all* n,
2. $f(i) = \lim_{n\to\infty} f_n(i)$ *and* $g(i) := \lim_{n\to\infty} g_n(i)$ *exists for each* i,
3. $|f_n(i)| \leq g_n(i)$ *for all* i *and* n,
4. $\lim_{n\to\infty} \sum_{i=1}^{\infty} g_n(i) = \sum_{i=1}^{\infty} g(i) < \infty$.

Then

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = \sum_{i=1}^{\infty} \lim_{n\to\infty} f_n(i) = \sum_{i=1}^{\infty} f(i).$$

(Often this proposition is used in the special case where $g_n = g$ for all n.)

Exercise 1.1. Prove Proposition 1.8. **Hint:** Let $a_{kn}^{+} := \max(a_{kn}, 0)$ and $a_{kn}^{-} = \max(-a_{kn}, 0)$ and observe that; $a_{kn} = a_{kn}^{+} - a_{kn}^{-}$ and $|a_{kn}^{+}| + |a_{kn}^{-}| = |a_{kn}|$. Now apply Proposition 1.7 with a_{kn} replaced by a_{kn}^{+} and a_{kn}^{-}.

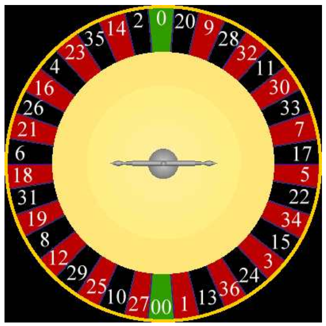
Exercise 1.2. Prove Proposition 1.10. **Hint:** apply the MCT by applying the monotone convergence theorem with $f_n(i) := \inf_{m\geq n} h_m(i)$.

Exercise 1.3. Prove Proposition 1.11. **Hint:** Apply Fatou's lemma twice. Once with $h_n(i) = g_n(i) + f_n(i)$ and once with $h_n(i) = g_n(i) - f_n(i)$.

2

Basic Probabilistic Notions

Definition 2.1. *A sample space Ω is a set which is to represents all possible outcomes of an "experiment."*



Example 2.2.

1. The sample space for flipping a coin one time could be taken to be, $\Omega = \{0, 1\}$.
2. The sample space for flipping a coin N -times could be taken to be, $\Omega = \{0, 1\}^N$ and for flipping an infinite number of times,
$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}}.$$
3. If we have a roulette wheel with 38 entries, then we might take
$$\Omega = \{00, 0, 1, 2, \dots, 36\}$$
for one spin,
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^N$$
for N spins, and
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^{\mathbb{N}}$$
for an infinite number of spins.
4. If we throw darts at a board of radius R, we may take
$$\Omega = D_R := \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le R\right\}$$
for one throw,
$$\Omega = D_R^N$$
for N throws, and
$$\Omega = D_R^{\mathbb{N}}$$
for an infinite number of throws.
5. Suppose we release a perfume particle at location $x \in \mathbb{R}^3$ and follow its motion for all time, $0 \le t < \infty$. In this case, we might take,
$$\Omega = \left\{\omega \in C\left([0, \infty), \mathbb{R}^3\right) : \omega(0) = x\right\}.$$

Definition 2.3. *An **event**, A, is a subset of Ω. Given $A \subset \Omega$ we also define the **indicator function of** A by*
$$1_A(\omega) := \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}.$$

Example 2.4. Suppose that $\Omega = \{0, 1\}^{\mathbb{N}}$ is the sample space for flipping a coin an infinite number of times. Here $\omega_n = 1$ represents the fact that a head was thrown on the n^{th} – toss, while $\omega_n = 0$ represents a tail on the n^{th} – toss.

1. $A = \{\omega \in \Omega : \omega_3 = 1\}$ represents the event that the third toss was a head.
2. $A = \cup_{i=1}^{\infty} \{\omega \in \Omega : \omega_i = \omega_{i+1} = 1\}$ represents the event that (at least) two heads are tossed twice in a row at some time.
3. $A = \cap_{N=1}^{\infty} \cup_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where there are infinitely many heads tossed in the sequence.
4. $A = \cup_{N=1}^{\infty} \cap_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where heads occurs from some time onwards, i.e. $\omega \in A$ iff there exists, $N = N(\omega)$ such that $\omega_n = 1$ for all $n \ge N$.

Ideally we would like to assign a probability, $P(A)$, to all events $A \subset \Omega$. Given a physical experiment, we think of assigning this probability as follows. Run the experiment many times to get sample points, $\omega(n) \in \Omega$ for each $n \in \mathbb{N}$, then try to "define" $P(A)$ by

$$P(A) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} 1_A(\omega(k)) \tag{2.1}$$
$$= \lim_{N \to \infty} \frac{1}{N} \#\{1 \le k \le N : \omega(k) \in A\}. \tag{2.2}$$

That is we think of $P(A)$ as being the long term relative frequency that the event A occurred for the sequence of experiments, $\{\omega(k)\}_{k=1}^{\infty}$.

Similarly supposed that A and B are two events and we wish to know how likely the event A is given that we know that B has occurred. Thus we would like to compute:

$$P(A|B) = \lim_{N\to\infty} \frac{\#\{k : 1 \le k \le N \text{ and } \omega_k \in A \cap B\}}{\#\{k : 1 \le k \le N \text{ and } \omega_k \in B\}},$$

which represents the frequency that A occurs given that we know that B has occurred. This may be rewritten as

$$\begin{aligned} P(A|B) &= \lim_{N\to\infty} \frac{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega_k \in A \cap B\}}{\frac{1}{N}\#\{k : 1 \le k \le N \text{ and } \omega_k \in B\}} \\ &= \frac{P(A\cap B)}{P(B)}. \end{aligned}$$

Definition 2.5. *If B is a non-null event, i.e. $P(B) > 0$, define the* ***conditional probability of** A **given** B by,*

$$P(A|B) := \frac{P(A\cap B)}{P(B)}.$$

There are of course a number of problems with this definition of P in Eq. (2.1) including the fact that it is not mathematical nor necessarily well defined. For example the limit may not exist. But ignoring these technicalities for the moment, let us point out three key properties that P should have.

1. $P(A) \in [0,1]$ for all $A \subset \Omega$.
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
3. **Additivity.** If A and B are disjoint event, i.e. $A \cap B = AB = \emptyset$, then $1_{A\cup B} = 1_A + 1_B$ so that

$$\begin{aligned} P(A\cup B) &= \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} 1_{A\cup B}(\omega(k)) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\left[1_A(\omega(k)) + 1_B(\omega(k))\right] \\ &= \lim_{N\to\infty}\left[\frac{1}{N}\sum_{k=1}^{N} 1_A(\omega(k)) + \frac{1}{N}\sum_{k=1}^{N} 1_B(\omega(k))\right] \\ &= P(A) + P(B). \end{aligned}$$

4. **Countable Additivity.** If $\{A_j\}_{j=1}^{\infty}$ are pairwise disjoint events (i.e. $A_j \cap A_k = \emptyset$ for all $j \ne k$), then again, $1_{\cup_{j=1}^{\infty} A_j} = \sum_{j=1}^{\infty} 1_{A_j}$ and therefore we might hope that,

$$\begin{aligned} P\left(\cup_{j=1}^{\infty} A_j\right) &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} 1_{\cup_{j=1}^{\infty} A_j}(\omega(k)) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{\infty} 1_{A_j}(\omega(k)) \\ &= \lim_{N\to\infty}\sum_{j=1}^{\infty}\frac{1}{N}\sum_{k=1}^{N} 1_{A_j}(\omega(k)) \\ &\overset{?}{=} \sum_{j=1}^{\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} 1_{A_j}(\omega(k)) \text{ (by a leap of faith)} \\ &= \sum_{j=1}^{\infty} P(A_j). \end{aligned}$$

Example 2.6. Let us consider the tossing of a coin N times with a fair coin. In this case we would expect that every $\omega \in \Omega$ is equally likely, i.e. $P(\{\omega\}) = \frac{1}{2^N}$. Assuming this we are then forced to define

$$P(A) = \frac{1}{2^N}\#(A).$$

Observe that this probability has the following property. Suppose that $\sigma \in \{0,1\}^k$ is a given sequence, then

$$P(\{\omega : (\omega_1,\dots,\omega_k) = \sigma\}) = \frac{1}{2^N}\cdot 2^{N-k} = \frac{1}{2^k}.$$

That is if we ignore the flips after time k, the resulting probabilities are the same as if we only flipped the coin k times.

Example 2.7. The previous example suggests that if we flip a fair coin an infinite number of times, so that now $\Omega = \{0,1\}^{\mathbb{N}}$, then we should define

$$P(\{\omega \in \Omega : (\omega_1,\dots,\omega_k) = \sigma\}) = \frac{1}{2^k} \tag{2.3}$$

for any $k \ge 1$ and $\sigma \in \{0,1\}^k$. Assuming there exists a probability, $P : 2^{\Omega} \to [0,1]$ such that Eq. (2.3) holds, we would like to compute, for example, the probability of the event B where an infinite number of heads are tossed. To try to compute this, let

$$\begin{aligned} A_n &= \{\omega \in \Omega : \omega_n = 1\} = \{\text{heads at time } n\} \\ B_N &:= \cup_{n\ge N} A_n = \{\text{at least one heads at time } N \text{ or later}\} \end{aligned}$$

and

$$B = \cap_{N=1}^{\infty} B_N = \{A_n \text{ i.o.}\} = \cap_{N=1}^{\infty}\cup_{n\ge N} A_n.$$

Since

$$B_N^c = \cap_{n\geq N} A_n^c \subset \cap_{M\geq n\geq N} A_n^c = \{\omega \in \Omega : \omega_N = \omega_{N+1} = \cdots = \omega_M = 0\},$$

we see that

$$P(B_N^c) \leq \frac{1}{2^{M-N}} \to 0 \text{ as } M \to \infty.$$

Therefore, $P(B_N) = 1$ for all N. If we assume that P is continuous under taking decreasing limits we may conclude, using $B_N \downarrow B$, that

$$P(B) = \lim_{N\to\infty} P(B_N) = 1.$$

Without this continuity assumption we would not be able to compute $P(B)$.

The unfortunate fact is that we can not always assign a desired probability function, $P(A)$, for all $A \subset \Omega$. For example we have the following negative theorem.

Theorem 2.8 (No-Go Theorem). *Let $S = \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle. Then there is no probability function, $P : 2^S \to [0,1]$ such that $P(S) = 1$, P is invariant under rotations, and P is continuous under taking decreasing limits.*

Proof. We are going to use the fact proved below in Proposition 5.3, that the continuity condition on P is equivalent to the σ – additivity of P. For $z \in S$ and $N \subset S$ let

$$zN := \{zn \in S : n \in N\}, \tag{2.4}$$

that is to say $e^{i\theta}N$ is the set N rotated counter clockwise by angle θ. By assumption, we are supposing that

$$P(zN) = P(N) \tag{2.5}$$

for all $z \in S$ and $N \subset S$.

Let

$$R := \{z = e^{i2\pi t} : t \in \mathbb{Q}\} = \{z = e^{i2\pi t} : t \in [0,1) \cap \mathbb{Q}\}$$

– a countable subgroup of S. As above R acts on S by rotations and divides S up into equivalence classes, where $z, w \in S$ are equivalent if $z = rw$ for some $r \in R$. Choose (using the axiom of choice) one representative point n from each of these equivalence classes and let $N \subset S$ be the set of these representative points. Then every point $z \in S$ may be uniquely written as $z = nr$ with $n \in N$ and $r \in R$. That is to say

$$S = \sum_{r\in R} (rN) \tag{2.6}$$

where $\sum_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (2.5) and (2.6),

$$1 = P(S) = \sum_{r\in R} P(rN) = \sum_{r\in R} P(N). \tag{2.7}$$

We have thus arrived at a contradiction, since the right side of Eq. (2.7) is either equal to 0 or to ∞ depending on whether $P(N) = 0$ or $P(N) > 0$. ∎

To avoid this problem, we are going to have to relinquish the idea that P should necessarily be defined on all of 2^Ω. So we are going to only define P on particular subsets, $\mathcal{B} \subset 2^\Omega$. We will developed this below.

Part II

Formal Development

3

Preliminaries

3.1 Set Operations

Let $\mathbb{N}$ denote the positive integers, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ be the non-negative integers and $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$ – the positive and negative integers including 0, $\mathbb{Q}$ the rational numbers, $\mathbb{R}$ the real numbers, and $\mathbb{C}$ the complex numbers. We will also use $\mathbb{F}$ to stand for either of the fields $\mathbb{R}$ or $\mathbb{C}$.

Notation 3.1 *Given two sets X and Y, let Y^X denote the collection of all functions $f : X \to Y$. If $X = \mathbb{N}$, we will say that $f \in Y^{\mathbb{N}}$ is a sequence with values in Y and often write f_n for $f(n)$ and express f as $\{f_n\}_{n=1}^\infty$. If $X = \{1, 2, \ldots, N\}$, we will write Y^N in place of $Y^{\{1,2,\ldots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \ldots, f_N)$ where $f_n = f(n)$.*

Notation 3.2 *More generally if $\{X_\alpha : \alpha \in A\}$ is a collection of non-empty sets, let $X_A = \prod_{\alpha \in A} X_\alpha$ and $\pi_\alpha : X_A \to X_\alpha$ be the canonical projection map defined by $\pi_\alpha(x) = x_\alpha$. If If $X_\alpha = X$ for some fixed space X, then we will write $\prod_{\alpha \in A} X_\alpha$ as X^A rather than X_A.*

Recall that an element $x \in X_A$ is a "**choice function**," i.e. an assignment $x_\alpha := x(\alpha) \in X_\alpha$ for each $\alpha \in A$. The **axiom of choice** states that $X_A \neq \emptyset$ provided that $X_\alpha \neq \emptyset$ for each $\alpha \in A$.

Notation 3.3 *Given a set X, let 2^X denote the **power set** of X – the collection of all subsets of X including the empty set.*

The reason for writing the power set of X as 2^X is that if we think of 2 meaning $\{0, 1\}$, then an element of $a \in 2^X = \{0, 1\}^X$ is completely determined by the set

$$A := \{x \in X : a(x) = 1\} \subset X.$$

In this way elements in $\{0, 1\}^X$ are in one to one correspondence with subsets of X.

For $A \in 2^X$ let

$$A^c := X \setminus A = \{x \in X : x \notin A\}$$

and more generally if $A, B \subset X$ let

$$B \setminus A := \{x \in B : x \notin A\} = B \cap A^c.$$

We also define the symmetric difference of A and B by

$$A \triangle B := (B \setminus A) \cup (A \setminus B).$$

As usual if $\{A_\alpha\}_{\alpha \in I}$ is an indexed collection of subsets of X we define the union and the intersection of this collection by

$$\cup_{\alpha \in I} A_\alpha := \{x \in X : \exists\, \alpha \in I \ni x \in A_\alpha\} \text{ and}$$
$$\cap_{\alpha \in I} A_\alpha := \{x \in X : x \in A_\alpha \,\forall\, \alpha \in I\}.$$

Notation 3.4 *We will also write $\sum_{\alpha \in I} A_\alpha$ for $\cup_{\alpha \in I} A_\alpha$ in the case that $\{A_\alpha\}_{\alpha \in I}$ are pairwise disjoint, i.e. $A_\alpha \cap A_\beta = \emptyset$ if $\alpha \neq \beta$.*

Notice that $\cup$ is closely related to $\exists$ and $\cap$ is closely related to $\forall$. For example let $\{A_n\}_{n=1}^\infty$ be a sequence of subsets from X and define

$$\inf_{k \geq n} A_n := \cap_{k \geq n} A_k,$$
$$\sup_{k \geq n} A_n := \cup_{k \geq n} A_k,$$
$$\limsup_{n \to \infty} A_n := \{A_n \text{ i.o.}\} := \{x \in X : \# \{n : x \in A_n\} = \infty\}$$

and

$$\liminf_{n \to \infty} A_n := \{A_n \text{ a.a.}\} := \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\}.$$

(One should read $\{A_n \text{ i.o.}\}$ as A_n infinitely often and $\{A_n \text{ a.a.}\}$ as A_n almost always.) Then $x \in \{A_n \text{ i.o.}\}$ iff

$$\forall N \in \mathbb{N}\ \exists\ n \geq N \ni x \in A_n$$

and this may be expressed as

$$\{A_n \text{ i.o.}\} = \cap_{N=1}^\infty \cup_{n \geq N} A_n.$$

Similarly, $x \in \{A_n \text{ a.a.}\}$ iff

$$\exists\ N \in \mathbb{N} \ni \forall\ n \geq N,\ \ x \in A_n$$

which may be written as

$$\{A_n \text{ a.a.}\} = \cup_{N=1}^\infty \cap_{n \geq N} A_n.$$

Definition 3.5. *Given a set $A \subset X$, let*

$$1_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

be the ***indicator function*** *of A.*

Lemma 3.6. *We have:*

1. $(\cup_n A_n)^c = \cap_n A_n^c$,
2. $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ a.a.}\}$,
3. $\limsup\limits_{n\to\infty} A_n = \{x \in X : \sum_{n=1}^\infty 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n\to\infty} A_n = \left\{x \in X : \sum_{n=1}^\infty 1_{A_n^c}(x) < \infty\right\}$,
5. $\sup_{k\geq n} 1_{A_k}(x) = 1_{\cup_{k\geq n} A_k} = 1_{\sup_{k\geq n} A_k}$,
6. $\inf_{k\geq n} 1_{A_k}(x) = 1_{\cap_{k\geq n} A_k} = 1_{\inf_{k\geq n} A_k}$,
7. $1_{\limsup\limits_{n\to\infty} A_n} = \limsup\limits_{n\to\infty} 1_{A_n}$, and
8. $1_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} 1_{A_n}$.

Definition 3.7. *A set X is said to be* ***countable*** *if is empty or there is an injective function $f : X \to \mathbb{N}$, otherwise X is said to be* ***uncountable.***

Lemma 3.8 (Basic Properties of Countable Sets).

1. *If $A \subset X$ is a subset of a countable set X then A is countable.*
2. *Any infinite subset $\Lambda \subset \mathbb{N}$ is in one to one correspondence with $\mathbb{N}$.*
3. *A non-empty set X is countable iff there exists a surjective map, $g : \mathbb{N} \to X$.*
4. *If X and Y are countable then $X \times Y$ is countable.*
5. *Suppose for each $m \in \mathbb{N}$ that A_m is a countable subset of a set X, then $A = \cup_{m=1}^\infty A_m$ is countable. In short, the countable union of countable sets is still countable.*
6. *If X is an infinite set and Y is a set with at least two elements, then Y^X is uncountable. In particular 2^X is uncountable for any infinite set X.*

Proof. 1. If $f : X \to \mathbb{N}$ is an injective map then so is the restriction, $f|_A$, of f to the subset A. 2. Let $f(1) = \min \Lambda$ and define f inductively by

$$f(n+1) = \min\left(\Lambda \setminus \{f(1), \dots, f(n)\}\right).$$

Since Λ is infinite the process continues indefinitely. The function $f : \mathbb{N} \to \Lambda$ defined this way is a bijection.

3. If $g : \mathbb{N} \to X$ is a surjective map, let

$$f(x) = \min g^{-1}(\{x\}) = \min\{n \in \mathbb{N} : f(n) = x\}.$$

Then $f : X \to \mathbb{N}$ is injective which combined with item

2. (taking $\Lambda = f(X)$) shows X is countable. Conversely if $f : X \to \mathbb{N}$ is injective let $x_0 \in X$ be a fixed point and define $g : \mathbb{N} \to X$ by $g(n) = f^{-1}(n)$ for $n \in f(X)$ and $g(n) = x_0$ otherwise.

4. Let us first construct a bijection, h, from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$. To do this put the elements of $\mathbb{N} \times \mathbb{N}$ into an array of the form

$$\begin{pmatrix} (1,1) & (1,2) & (1,3) & \dots \\ (2,1) & (2,2) & (2,3) & \dots \\ (3,1) & (3,2) & (3,3) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and then "count" these elements by counting the sets $\{(i,j) : i+j = k\}$ one at a time. For example let $h(1) = (1,1)$, $h(2) = (2,1)$, $h(3) = (1,2)$, $h(4) = (3,1)$, $h(5) = (2,2)$, $h(6) = (1,3)$ and so on. If $f : \mathbb{N} \to X$ and $g : \mathbb{N} \to Y$ are surjective functions, then the function $(f \times g) \circ h : \mathbb{N} \to X \times Y$ is surjective where $(f \times g)(m,n) := (f(m), g(n))$ for all $(m,n) \in \mathbb{N} \times \mathbb{N}$.

5. If $A = \emptyset$ then A is countable by definition so we may assume $A \neq \emptyset$. With out loss of generality we may assume $A_1 \neq \emptyset$ and by replacing A_m by A_1 if necessary we may also assume $A_m \neq \emptyset$ for all m. For each $m \in \mathbb{N}$ let $a_m : \mathbb{N} \to A_m$ be a surjective function and then define $f : \mathbb{N} \times \mathbb{N} \to \cup_{m=1}^\infty A_m$ by $f(m,n) := a_m(n)$. The function f is surjective and hence so is the composition, $f \circ h : \mathbb{N} \to \cup_{m=1}^\infty A_m$, where $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ is the bijection defined above.

6. Let us begin by showing $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$ is uncountable. For sake of contradiction suppose $f : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$ is a surjection and write $f(n)$ as $(f_1(n), f_2(n), f_3(n), \dots)$. Now define $a \in \{0,1\}^{\mathbb{N}}$ by $a_n := 1 - f_n(n)$. By construction $f_n(n) \neq a_n$ for all n and so $a \notin f(\mathbb{N})$. This contradicts the assumption that f is surjective and shows $2^{\mathbb{N}}$ is uncountable. For the general case, since $Y_0^X \subset Y^X$ for any subset $Y_0 \subset Y$, if Y_0^X is uncountable then so is Y^X. In this way we may assume Y_0 is a two point set which may as well be $Y_0 = \{0,1\}$. Moreover, since X is an infinite set we may find an injective map $x : \mathbb{N} \to X$ and use this to set up an injection, $i : 2^{\mathbb{N}} \to 2^X$ by setting $i(A) := \{x_n : n \in \mathbb{N}\} \subset X$ for all $A \subset \mathbb{N}$. If 2^X were countable we could find a surjective map $f : 2^X \to \mathbb{N}$ in which case $f \circ i : 2^{\mathbb{N}} \to \mathbb{N}$ would be surjective as well. However this is impossible since we have already seed that $2^{\mathbb{N}}$ is uncountable. ■

3.2 Exercises

Let $f : X \to Y$ be a function and $\{A_i\}_{i\in I}$ be an indexed family of subsets of Y, verify the following assertions.

Exercise 3.1. $(\cap_{i\in I} A_i)^c = \cup_{i\in I} A_i^c$.

Exercise 3.2. Suppose that $B \subset Y$, show that $B \setminus (\cup_{i\in I} A_i) = \cap_{i\in I}(B \setminus A_i)$.

Exercise 3.3. $f^{-1}(\cup_{i\in I} A_i) = \cup_{i\in I} f^{-1}(A_i)$.

Exercise 3.4. $f^{-1}(\cap_{i\in I} A_i) = \cap_{i\in I} f^{-1}(A_i)$.

Exercise 3.5. Find a counterexample which shows that $f(C \cap D) = f(C) \cap f(D)$ need not hold.

Example 3.9. Let $X = \{a,b,c\}$ and $Y = \{1,2\}$ and define $f(a) = f(b) = 1$ and $f(c) = 2$. Then $\emptyset = f(\{a\} \cap \{b\}) \neq f(\{a\}) \cap f(\{b\}) = \{1\}$ and $\{1,2\} = f(\{a\}^c) \neq f(\{a\})^c = \{2\}$.

3.3 Algebraic sub-structures of sets

Definition 3.10. *A collection of subsets $\mathcal{A}$ of a set X is a π – **system** or **multiplicative system** if $\mathcal{A}$ is closed under taking finite intersections.*

Definition 3.11. *A collection of subsets $\mathcal{A}$ of a set X is an **algebra (Field)** if*

1. *$\emptyset, X \in \mathcal{A}$*
2. *$A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$*
3. *$\mathcal{A}$ is closed under finite unions, i.e. if $A_1, \dots, A_n \in \mathcal{A}$ then $A_1 \cup \dots \cup A_n \in \mathcal{A}$. In view of conditions 1. and 2., 3. is equivalent to*

3′. *$\mathcal{A}$ is closed under finite intersections.*

Definition 3.12. *A collection of subsets $\mathcal{B}$ of X is a σ – **algebra** (or sometimes called a σ – **field**) if $\mathcal{B}$ is an algebra which also closed under countable unions, i.e. if $\{A_i\}_{i=1}^\infty \subset \mathcal{B}$, then $\cup_{i=1}^\infty A_i \in \mathcal{B}$. (Notice that since $\mathcal{B}$ is also closed under taking complements, $\mathcal{B}$ is also closed under taking countable intersections.)*

Example 3.13. Here are some examples of algebras.

1. $\mathcal{B} = 2^X$, then $\mathcal{B}$ is a σ – algebra.
2. $\mathcal{B} = \{\emptyset, X\}$ is a σ – algebra called the trivial σ – field.
3. Let $X = \{1,2,3\}$, then $\mathcal{A} = \{\emptyset, X, \{1\}, \{2,3\}\}$ is an algebra while, $\mathcal{S} := \{\emptyset, X, \{2,3\}\}$ is a not an algebra but is a π – system.

Proposition 3.14. *Let $\mathcal{E}$ be any collection of subsets of X. Then there exists a unique smallest algebra $\mathcal{A}(\mathcal{E})$ and σ – algebra $\sigma(\mathcal{E})$ which contains $\mathcal{E}$.*

Proof. Simply take

$$\mathcal{A}(\mathcal{E}) := \bigcap \{\mathcal{A} : \mathcal{A} \text{ is an algebra such that } \mathcal{E} \subset \mathcal{A}\}$$

and

$$\sigma(\mathcal{E}) := \bigcap \{\mathcal{M} : \mathcal{M} \text{ is a } \sigma - \text{algebra such that } \mathcal{E} \subset \mathcal{M}\}.$$

■

Example 3.15. Suppose $X = \{1,2,3\}$ and $\mathcal{E} = \{\emptyset, X, \{1,2\}, \{1,3\}\}$, see Figure 3.1. Then

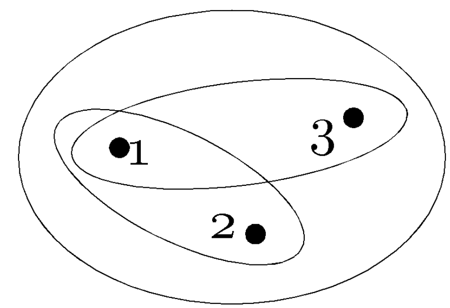


Fig. 3.1. A collection of subsets.

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = 2^X.$$

On the other hand if $\mathcal{E} = \{\{1,2\}\}$, then $\mathcal{A}(\mathcal{E}) = \{\emptyset, X, \{1,2\}, \{3\}\}$.

Exercise 3.6. Suppose that $\mathcal{E}_i \subset 2^X$ for $i = 1,2$. Show that $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \mathcal{A}(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \mathcal{A}(\mathcal{E}_1)$. Similarly show, $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$ iff $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$ and $\mathcal{E}_2 \subset \sigma(\mathcal{E}_1)$. Give a simple example where $\mathcal{A}(\mathcal{E}_1) = \mathcal{A}(\mathcal{E}_2)$ while $\mathcal{E}_1 \neq \mathcal{E}_2$.

In this course we will often be interested in the Borel σ – algebra on a topological space.

Definition 3.16 (Borel σ – field). *The **Borel** σ – **algebra**, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$, on $\mathbb{R}$ is the smallest σ -field containing all of the open subsets of $\mathbb{R}$. More generally if (X, τ) is a topological space, the Borel σ – algebra on X is $\mathcal{B}_X := \sigma(\tau)$ – i.e. the smallest σ – algebra containing all open (closed) subsets of X.*

Exercise 3.7. Verify the Borel σ – algebra, $\mathcal{B}_{\mathbb{R}}$, is generated by any of the following collection of sets:

$$1.\ \{(a,\infty) : a \in \mathbb{R}\},\ 2.\ \{(a,\infty) : a \in \mathbb{Q}\} \text{ or } 3.\ \{[a,\infty) : a \in \mathbb{Q}\}.$$

Hint: make use of Exercise 3.6.

We will postpone a more in depth study of σ – algebras until later. For now, let us concentrate on understanding the the simpler notion of an algebra.

Definition 3.17. *Let X be a set. We say that a family of sets $\mathcal{F} \subset 2^X$ is a **partition** of X if distinct members of $\mathcal{F}$ are disjoint and if X is the union of the sets in $\mathcal{F}$.*

Example 3.18. Let X be a set and $\mathcal{E} = \{A_1, \dots, A_n\}$ where $A_1, \dots, A_n$ is a partition of X. In this case

$$\mathcal{A}(\mathcal{E}) = \sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset \{1, 2, \dots, n\}\}$$

where $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. Notice that

$$\#\left(\mathcal{A}(\mathcal{E})\right) = \#(2^{\{1,2,\dots,n\}}) = 2^n.$$

Example 3.19. Suppose that X is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#\left(\mathcal{A}\right) < \infty$. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{A} : x \in A\} \in \mathcal{A},$$

wherein we have used $\mathcal{A}$ is finite to insure $A_x \in \mathcal{A}$. Hence A_x is the smallest set in $\mathcal{A}$ which contains x.

Now suppose that $y \in X$. If $x \in A_y$ then $A_x \subset A_y$ so that $A_x \cap A_y = A_x$. On the other hand, if $x \notin A_y$ then $x \in A_x \setminus A_y$ and therefore $A_x \subset A_x \setminus A_y$, i.e. $A_x \cap A_y = \emptyset$. Therefore we have shown, either $A_x \cap A_y = \emptyset$ or $A_x \cap A_y = A_x$. By reversing the roles of x and y it also follows that either $A_y \cap A_x = \emptyset$ or $A_y \cap A_x = A_y$. Therefore we may conclude, either $A_x = A_y$ or $A_x \cap A_y = \emptyset$ for all $x, y \in X$.

Let us now define $\{B_i\}_{i=1}^k$ to be an enumeration of $\{A_x\}_{x\in X}$. It is a straightforward to conclude that

$$\mathcal{A} = \{\cup_{i\in\Lambda} B_i : \Lambda \subset \{1, 2, \dots, k\}\}.$$

For example observe that for any $A \in \mathcal{A}$, we have $A = \cup_{x\in A} A_x = \cup_{i\in\Lambda} B_i$ where $\Lambda := \{i : B_i \subset A\}$.

Proposition 3.20. *Suppose that $\mathcal{B} \subset 2^X$ is a σ – algebra and $\mathcal{B}$ is at most a countable set. Then there exists a unique **finite** partition $\mathcal{F}$ of X such that $\mathcal{F} \subset \mathcal{B}$ and every element $B \in \mathcal{B}$ is of the form*

$$B = \cup\{A \in \mathcal{F} : A \subset B\}. \tag{3.1}$$

In particular $\mathcal{B}$ is actually a finite set and $\#\left(\mathcal{B}\right) = 2^n$ for some $n \in \mathbb{N}$.

Proof. We proceed as in Example 3.19. For each $x \in X$ let

$$A_x = \cap\{A \in \mathcal{B} : x \in A\} \in \mathcal{B},$$

wherein we have used $\mathcal{B}$ is a countable σ – algebra to insure $A_x \in \mathcal{B}$. Just as above either $A_x \cap A_y = \emptyset$ or $A_x = A_y$ and therefore $\mathcal{F} = \{A_x : x \in X\} \subset \mathcal{B}$ is a (necessarily countable) partition of X for which Eq. (3.1) holds for all $B \in \mathcal{B}$.

Enumerate the elements of $\mathcal{F}$ as $\mathcal{F} = \{P_n\}_{n=1}^N$ where $N \in \mathbb{N}$ or $N = \infty$. If $N = \infty$, then the correspondence

$$a \in \{0,1\}^{\mathbb{N}} \to A_a = \cup\{P_n : a_n = 1\} \in \mathcal{B}$$

is bijective and therefore, by Lemma 3.8, $\mathcal{B}$ is uncountable. Thus any countable σ – algebra is necessarily finite. This finishes the proof modulo the uniqueness assertion which is left as an exercise to the reader. ∎

Example 3.21 (Countable/Co-countable σ – Field). Let $X = \mathbb{R}$ and $\mathcal{E} := \{\{x\} : x \in \mathbb{R}\}$. Then $\sigma(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that A is countable or A^c is countable. Similarly, $\mathcal{A}(\mathcal{E})$ consists of those subsets, $A \subset \mathbb{R}$, such that A is finite or A^c is finite. More generally we have the following exercise.

Exercise 3.8. Let X be a set, I be an **infinite** index set, and $\mathcal{E} = \{A_i\}_{i\in I}$ be a partition of X. Prove the algebra, $\mathcal{A}(\mathcal{E})$, and that σ – algebra, $\sigma(\mathcal{E})$, generated by $\mathcal{E}$ are given by

$$\mathcal{A}(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \#(\Lambda) < \infty \text{ or } \#(\Lambda^c) < \infty\}$$

and

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I \text{ with } \Lambda \text{ countable or } \Lambda^c \text{ countable}\}$$

respectively. Here we are using the convention that $\cup_{i\in\Lambda} A_i := \emptyset$ when $\Lambda = \emptyset$. In particular if I is countable, then

$$\sigma(\mathcal{E}) = \{\cup_{i\in\Lambda} A_i : \Lambda \subset I\}.$$

Proposition 3.22. *Let X be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

$$\mathcal{A}(\mathcal{E}) := \{\text{finite unions of finite intersections of elements from } \mathcal{E}_c\}. \tag{3.2}$$

Proof. Let $\mathcal{A}$ denote the right member of Eq. (3.2). From the definition of an algebra, it is clear that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{E})$. Hence to finish that proof it suffices to show $\mathcal{A}$ is an algebra. The proof of these assertions are routine except for possibly showing that $\mathcal{A}$ is closed under complementation. To check $\mathcal{A}$ is closed under complementation, let $Z \in \mathcal{A}$ be expressed as

$$Z = \bigcup_{i=1}^{N} \bigcap_{j=1}^{K} A_{ij}$$

where $A_{ij} \in \mathcal{E}_c$. Therefore, writing $B_{ij} = A_{ij}^c \in \mathcal{E}_c$, we find that

$$Z^c = \bigcap_{i=1}^{N} \bigcup_{j=1}^{K} B_{ij} = \bigcup_{j_1,\dots,j_N=1}^{K} (B_{1j_1} \cap B_{2j_2} \cap \cdots \cap B_{Nj_N}) \in \mathcal{A}$$

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \cdots \cap B_{Nj_N}$ is a finite intersection of sets from $\mathcal{E}_c$. ■

Remark 3.23. One might think that in general $\sigma(\mathcal{E})$ may be described as the countable unions of countable intersections of sets in $\mathcal{E}^c$. However this is in general **false,** since if

$$Z = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{\infty} A_{ij}$$

with $A_{ij} \in \mathcal{E}_c$, then

$$Z^c = \bigcup_{j_1=1, j_2=1, \dots j_N=1, \dots}^{\infty} \left(\bigcap_{\ell=1}^{\infty} A_{\ell, j_\ell}^c \right)$$

which is now an **uncountable** union. Thus the above description is not correct. In general it is complicated to explicitly describe $\sigma(\mathcal{E})$, see Proposition 1.23 on page 39 of Folland for details. Also see Proposition 3.20.

Exercise 3.9. Let τ be a topology on a set X and $\mathcal{A} = \mathcal{A}(\tau)$ be the algebra generated by τ. Show $\mathcal{A}$ is the collection of subsets of X which may be written as finite union of sets of the form $F \cap V$ where F is closed and V is open.

Solution to Exercise (3.9). In this case τ_c is the collection of sets which are either open or closed. Now if $V_i \subset_o X$ and $F_j \sqsubset X$ for each j, then $(\cap_{i=1}^n V_i) \cap \left(\cap_{j=1}^m F_j\right)$ is simply a set of the form $V \cap F$ where $V \subset_o X$ and $F \sqsubset X$. Therefore the result is an immediate consequence of Proposition 3.22.

Definition 3.24. *A set $\mathcal{S} \subset 2^X$ is said to be an* ***semialgebra or elementary class*** *provided that*

- $\emptyset \in \mathcal{S}$
- *$\mathcal{S}$ is closed under finite intersections*
- *if $E \in \mathcal{S}$, then E^c is a finite disjoint union of sets from $\mathcal{S}$. (In particular $X = \emptyset^c$ is a finite disjoint union of elements from $\mathcal{S}$.)*

Proposition 3.25. *Suppose $\mathcal{S} \subset 2^X$ is a semi-field, then $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$.*

Proof. (Although it is possible to give a proof using Proposition 3.22, it is just as simple to give a direct proof.) Let $\mathcal{A}$ denote the collection of sets which may be written as finite disjoint unions of sets from $\mathcal{S}$. Clearly $\mathcal{S} \subset \mathcal{A} \subset \mathcal{A}(\mathcal{S})$ so it suffices to show $\mathcal{A}$ is an algebra since $\mathcal{A}(\mathcal{S})$ is the smallest algebra containing $\mathcal{S}$. By the properties of $\mathcal{S}$, we know that $\emptyset, X \in \mathcal{A}$. The following two steps now finish the proof.

1. ($\mathcal{A}$ is closed under finite intersections.) Suppose that $A_i = \sum_{F \in \Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \dots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{S}$. Then

$$\bigcap_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \left(\sum_{F \in \Lambda_i} F \right) = \bigcup_{(F_1, ,\dots, F_n) \in \Lambda_1 \times \cdots \times \Lambda_n} (F_1 \cap F_2 \cap \cdots \cap F_n)$$

and this is a disjoint (you check) union of elements from $\mathcal{S}$. Therefore $\mathcal{A}$ is closed under finite intersections.

2. ($\mathcal{A}$ is closed under complementation.) If $A = \sum_{F \in \Lambda} F$ with Λ being a finite collection of disjoint sets from $\mathcal{S}$, then $A^c = \bigcap_{F \in \Lambda} F^c$. Since, by assumption, $F^c \in \mathcal{A}$ for all $F \in \Lambda \subset \mathcal{S}$ and $\mathcal{A}$ is closed under finite intersections by step 1., it follows that $A^c \in \mathcal{A}$. ■

Example 3.26. Let $X = \mathbb{R}$, then

$$\begin{aligned} \mathcal{S} &:= \left\{ (a, b] \cap \mathbb{R} : a, b \in \bar{\mathbb{R}} \right\} \\ &= \{ (a, b] : a \in [-\infty, \infty) \text{ and } a < b < \infty \} \cup \{ \emptyset, \mathbb{R} \} \end{aligned}$$

is a semi-field. The algebra, $\mathcal{A}(\mathcal{S})$, generated by $\mathcal{S}$ consists of finite disjoint unions of sets from $\mathcal{S}$. For example,

$$A = (0, \pi] \cup (2\pi, 7] \cup (11, \infty) \in \mathcal{A}(\mathcal{S}).$$

Exercise 3.10. Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be semi-fields. Show the collection

$$\mathcal{S} := \{ A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B} \}$$

is also a semi-field.

Solution to Exercise (3.10). Clearly $\emptyset = \emptyset \times \emptyset \in \mathcal{E} = \mathcal{A} \times \mathcal{B}$. Let $A_i \in \mathcal{A}$ and $B_i \in \mathcal{B}$, then

$$\cap_{i=1}^n (A_i \times B_i) = (\cap_{i=1}^n A_i) \times (\cap_{i=1}^n B_i) \in \mathcal{A} \times \mathcal{B}$$

showing $\mathcal{E}$ is closed under finite intersections. For $A \times B \in \mathcal{E}$,

$$(A \times B)^c = (A^c \times B^c) \sum (A^c \times B) \sum (A \times B^c)$$

and by assumption $A^c = \sum_{i=1}^n A_i$ with $A_i \in \mathcal{A}$ and $B^c = \sum_{j=1}^m B_i$ with $B_j \in \mathcal{B}$. Therefore

$$A^c \times B^c = \left(\sum_{i=1}^n A_i \right) \times \left(\sum_{j=1}^m B_i \right) = \sum_{i=1,j=1}^{n,m} A_i \times B_i,$$

$$A^c \times B = \sum_{i=1}^n A_i \times B, \text{ and } A \times B^c = \sum_{j=1}^m A \times B_i$$

showing $(A \times B)^c$ may be written as finite disjoint union of elements from $\mathcal{S}$.

4

Finitely Additive Measures / Integration

Definition 4.1. *Suppose that $\mathcal{E} \subset 2^X$ is a collection of subsets of X and $\mu : \mathcal{E} \to [0,\infty]$ is a function. Then*

1. *μ is **additive or finitely additive on** $\mathcal{E}$ if*
$$\mu(E) = \sum_{i=1}^{n} \mu(E_i) \tag{4.1}$$
whenever $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ with $E_i \in \mathcal{E}$ for $i = 1, 2, \ldots, n < \infty$.
2. *μ is σ – **additive (or countable additive) on** $\mathcal{E}$ if Eq. (4.1) holds even when $n = \infty$.*
3. *μ is **sub-additive (finitely sub-additive) on** $\mathcal{E}$ if*
$$\mu(E) \le \sum_{i=1}^{n} \mu(E_i)$$
whenever $E = \bigcup_{i=1}^{n} E_i \in \mathcal{E}$ with $n \in \mathbb{N} \cup \{\infty\}$ ($n \in \mathbb{N}$).
4. *μ is a **finitely additive measure** if $\mathcal{E} = \mathcal{A}$ is an algebra, $\mu(\emptyset) = 0$, and μ is finitely additive on $\mathcal{A}$.*
5. *μ is a **premeasure** if μ is a finitely additive measure which is σ – additive on $\mathcal{A}$.*
6. *μ is a **measure** if μ is a premeasure on a σ – algebra. Furthermore if $\mu(X) = 1$, we say μ is a **probability measure** on X.*

Proposition 4.2 (Basic properties of finitely additive measures). *Suppose μ is a finitely additive measure on an algebra, $\mathcal{A} \subset 2^X$, $A, B \in \mathcal{A}$ with $A \subset B$and $\{A_j\}_{j=1}^n \subset \mathcal{A}$, then :*

1. *(μ is **monotone**) $\mu(A) \le \mu(B)$ if $A \subset B$.*
2. *For $A, B \in \mathcal{A}$, the following **strong additivity formula** holds;*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{4.2}$$
3. *(μ is **finitely subbadditive**) $\mu(\cup_{j=1}^n A_j) \le \sum_{j=1}^n \mu(A_j)$.*
4. *μ is sub-additive on $\mathcal{A}$ iff*
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \sum_{i=1}^{\infty} A_i \tag{4.3}$$
where $A \in \mathcal{A}$ and $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ are pairwise disjoint sets.
5. *(μ is **countably superadditive**) If $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then*
$$\mu\left(\sum_{i=1}^{\infty} A_i\right) \ge \sum_{i=1}^{\infty} \mu(A_i). \tag{4.4}$$
(See Remark 4.9 for example where this inequality is strict.)
6. *A finitely additive measure, μ, is a premeasure iff μ is subadditve.*

Proof.

1. Since B is the disjoint union of A and $(B \setminus A)$ and $B \setminus A = B \cap A^c \in \mathcal{A}$ it follows that
$$\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A).$$
2. Since
$$A \cup B = [A \setminus (A \cap B)] \sum [B \setminus (A \cap B)] \sum A \cap B,$$
$$\begin{aligned}\mu(A \cup B) &= \mu(A \cup B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B).\end{aligned}$$
Adding $\mu(A \cap B)$ to both sides of this equation proves Eq. (4.2).
3. Let $\widetilde{E}_j = E_j \setminus (E_1 \cup \cdots \cup E_{j-1})$ so that the $\tilde{E}_j$'s are pair-wise disjoint and $E = \cup_{j=1}^n \widetilde{E}_j$. Since $\tilde{E}_j \subset E_j$ it follows from the monotonicity of μ that
$$\mu(E) = \sum_{j=1}^{n} \mu(\widetilde{E}_j) \le \sum_{j=1}^{n} \mu(E_j).$$
4. If $A = \bigcup_{i=1}^{\infty} B_i$ with $A \in \mathcal{A}$ and $B_i \in \mathcal{A}$, then $A = \sum_{i=1}^{\infty} A_i$ where $A_i := B_i \setminus (B_1 \cup \ldots B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of μ and Eq. (4.3)
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \le \sum_{i=1}^{\infty} \mu(B_i).$$
5. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then $\sum_{i=1}^{n} A_i \subset A$ for all n and so by the monotonicity and finite additivity of μ, $\sum_{i=1}^{n} \mu(A_i) \le \mu(A)$. Letting $n \to \infty$ in this equation shows μ is superadditive.
6. This is a combination of items 5. and 6.

■

4.1 Examples of Measures

Most $\sigma-$ algebras and σ -additive measures are somewhat difficult to describe and define. However, there are a few special cases where we can describe explicitly what is going on.

Example 4.3. Suppose that Ω is a finite set, $\mathcal{B} := 2^{\Omega}$, and $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1.$$

Then

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \subset \Omega$$

defines a measure on 2^{Ω}.

Example 4.4. Suppose that X is any set and $x \in X$ is a point. For $A \subset X$, let

$$\delta_x(A) = \begin{cases} 1 \text{ if } & x \in A \\ 0 \text{ if } & x \notin A. \end{cases}$$

Then $\mu = \delta_x$ is a measure on X called the Dirac delta measure at x.

Example 4.5. Suppose $\mathcal{B} \subset 2^X$ is a σ algebra, μ is a measure on $\mathcal{B}$, and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on $\mathcal{B}$. Moreover, if J is an index set and $\{\mu_j\}_{j\in J}$ are all measures on $\mathcal{B}$, then $\mu = \sum_{j=1}^{\infty} \mu_j$, i.e.

$$\mu(A) := \sum_{j=1}^{\infty} \mu_j(A) \text{ for all } A \in \mathcal{B},$$

defines another measure on $\mathcal{B}$. To prove this we must show that μ is countably additive. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i \in \mathcal{B}$, then (using Tonelli for sums, Proposition 1.7),

$$\begin{aligned} \mu(A) &= \sum_{j=1}^{\infty} \mu_j(A) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \mu_j(A_i) \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu_j(A_i) = \sum_{i=1}^{\infty} \mu(A_i). \end{aligned}$$

Example 4.6. Suppose that X is a countable set and $\lambda : X \to [0,\infty]$ is a function. Let $X = \{x_n\}_{n=1}^{\infty}$ be an enumeration of X and then we may define a measure μ on 2^X by,

$$\mu = \mu_\lambda := \sum_{n=1}^{\infty} \lambda(x_n)\delta_{x_n}.$$

We will now show this measure is independent of our choice of enumeration of X by showing,

$$\mu(A) = \sum_{x\in A} \lambda(x) := \sup_{\Lambda \subset\subset A} \sum_{x\in\Lambda} \lambda(x) \ \forall\ A \subset X. \tag{4.5}$$

Here we are using the notation, $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.

To verify Eq. (4.5), let $M := \sup_{\Lambda\subset\subset A} \sum_{x\in\Lambda} \lambda(x)$ and for each $N \in \mathbb{N}$ let

$$\Lambda_N := \{x_n : x_n \in A \text{ and } 1 \le n \le N\}.$$

Then by definition of μ,

$$\begin{aligned} \mu(A) &= \sum_{n=1}^{\infty} \lambda(x_n)\delta_{x_n}(A) = \lim_{N\to\infty} \sum_{n=1}^{N} \lambda(x_n) 1_{x_n\in A} \\ &= \lim_{N\to\infty} \sum_{x\in\Lambda_N} \lambda(x) \le M. \end{aligned}$$

On the other hand if $\Lambda \subset\subset A$, then

$$\sum_{x\in\Lambda} \lambda(x) = \sum_{n:\ x_n\in\Lambda} \lambda(x_n) = \mu(\Lambda) \le \mu(A)$$

from which it follows that $M \le \mu(A)$. This shows that μ is independent of how we enumerate X.

The above example has a natural extension to the case where X is uncountable and $\lambda : X \to [0,\infty]$ is any function. In this setting we simply may define $\mu : 2^X \to [0,\infty]$ using Eq. (4.5). We leave it to the reader to verify that this is indeed a measure on 2^X.

We will construct many more measure in Chapter 5 below. The starting point of these constructions will be the construction of finitely additive measures using the next proposition.

Proposition 4.7 (Construction of Finitely Additive Measures). *Suppose $\mathcal{S} \subset 2^X$ is a semi-algebra (see Definition 3.24) and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ is the algebra generated by $\mathcal{S}$. Then every additive function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure (which we still denote by μ) on $\mathcal{A}$.*

Proof. Since (by Proposition 3.25) every element $A \in \mathcal{A}$ is of the form $A = \sum_i E_i$ for a finite collection of $E_i \in \mathcal{S}$, it is clear that if μ extends to a measure then the extension is unique and must be given by

$$\mu(A) = \sum_i \mu(E_i). \tag{4.6}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (4.6) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

$$\sum_i \mu(E_i) = \sum_j \mu(F_j). \tag{4.7}$$

But $E_i = \sum_j (E_i \cap F_j)$ and the additivity of μ on $\mathcal{S}$ implies $\mu(E_i) = \sum_j \mu(E_i \cap F_j)$ and hence

$$\sum_i \mu(E_i) = \sum_i \sum_j \mu(E_i \cap F_j) = \sum_{i,j} \mu(E_i \cap F_j).$$

Similarly,

$$\sum_j \mu(F_j) = \sum_{i,j} \mu(E_i \cap F_j)$$

which combined with the previous equation shows that Eq. (4.7) holds. It is now easy to verify that μ extended to $\mathcal{A}$ as in Eq. (4.6) is an additive measure on $\mathcal{A}$. ■

Proposition 4.8. *Let $X = \mathbb{R}$, $\mathcal{S}$ be the semi-algebra,*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{4.8}$$

and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{S}$, see Proposition 3.25. To each finitely additive probability measures $\mu : \mathcal{A} \to [0, \infty]$, there is a unique increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ and

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) \ \forall \ a \le b \text{ in } \bar{\mathbb{R}}. \tag{4.9}$$

Conversely, given an increasing function $F : \bar{\mathbb{R}} \to [0, 1]$ such that $F(-\infty) = 0$, $F(\infty) = 1$ there is a unique finitely additive measure $\mu = \mu_F$ on $\mathcal{A}$ such that the relation in Eq. (4.9) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a \downarrow -\infty} F(a)$ and $F(\infty) = \lim_{b \uparrow \infty} F(b)$.)

Proof. Given a finitely additive probability measure μ, let

$$F(x) := \mu((-\infty, x] \cap \mathbb{R}) \text{ for all } x \in \bar{\mathbb{R}}.$$

Then $F(\infty) = 1$, $F(-\infty) = 0$ and for $b > a$,

$$F(b) - F(a) = \mu((-\infty, b] \cap \mathbb{R}) - \mu((-\infty, a]) = \mu((a, b] \cap \mathbb{R}).$$

Conversely, suppose $F : \bar{\mathbb{R}} \to [0, 1]$ as in the statement of the theorem is given. Define μ on $\mathcal{S}$ using the formula in Eq. (4.9). The argument will be completed by showing μ is additive on $\mathcal{S}$ and hence, by Proposition 4.7, has a unique extension to a finitely additive measure on $\mathcal{A}$. Suppose that

$$(a, b] = \sum_{i=1}^n (a_i, b_i].$$

By reordering $(a_i, b_i]$ if necessary, we may assume that

$$a = a_1 < b_1 = a_2 < b_2 = a_3 < \cdots < b_{n-1} = a_n < b_n = b.$$

Therefore, by the telescoping series argument,

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) = \sum_{i=1}^n [F(b_i) - F(a_i)] = \sum_{i=1}^n \mu((a_i, b_i] \cap \mathbb{R}).$$

■

Remark 4.9. Suppose that $F : \bar{\mathbb{R}} \to \bar{\mathbb{R}}$ is any non-decreasing function such that $F(\mathbb{R}) \subset \mathbb{R}$. Then the same methods used in the proof of Proposition 4.8 shows that there exists a unique finitely additive measure, $\mu = \mu_F$, on $\mathcal{A} = \mathcal{A}(\mathcal{S})$ such that Eq. (4.9) holds. If $F(\infty) > \lim_{b \uparrow \infty} F(b)$ and $A_i = (i, i + 1]$ for $i \in \mathbb{N}$, then

$$\begin{aligned}
\sum_{i=1}^\infty \mu_F(A_i) &= \sum_{i=1}^\infty (F(i+1) - F(i)) = \lim_{N \to \infty} \sum_{i=1}^N (F(i+1) - F(i)) \\
&= \lim_{N \to \infty} (F(N+1) - F(1)) < F(\infty) - F(1) = \mu_F(\cup_{i=1}^\infty A_i).
\end{aligned}$$

This shows that strict inequality can hold in Eq. (4.4) and that μ_F is **not** a premeasure. Similarly one shows μ_F is **not** a premeasure if $F(-\infty) < \lim_{a \downarrow -\infty} F(a)$ or if F is **not** right continuous at some point $a \in \mathbb{R}$. Indeed, in the latter case consider

$$(a, a+1] = \sum_{n=1}^\infty (a + \frac{1}{n+1}, a + \frac{1}{n}].$$

Working as above we find,

$$\sum_{n=1}^\infty \mu_F\left((a + \frac{1}{n+1}, a + \frac{1}{n}]\right) = F(a+1) - F(a+)$$

while $\mu_F((a, a+1]) = F(a+1) - F(a)$. We will eventually show in Chapter 5 below that μ_F extends uniquely to a $\sigma -$ additive measure on $\mathcal{B}_{\mathbb{R}}$ whenever F is increasing, right continuous, and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$.

Before constructing σ – additive measures (see Chapter 5 below), we are going to pause to discuss a preliminary notion of integration and develop some of its properties. Hopefully this will help the reader to develop the necessary intuition before heading to the general theory. First we need to describe the functions we are allowed to integrate.

4.2 Simple Random Variables

Definition 4.10 (Simple random variables). *A function, $f : \Omega \to Y$ is said to be* ***simple*** *if $f(\Omega) \subset Y$ is a finite set. If $\mathcal{A} \subset 2^{\Omega}$ is an algebra, we say that a simple function $f : \Omega \to Y$ is* ***measurable*** *if $\{f = y\} := f^{-1}(\{y\}) \in \mathcal{A}$ for all $y \in Y$. A measurable simple function, $f : \Omega \to \mathbb{C}$, is called a* ***simple random variable*** *relative to $\mathcal{A}$.*

Notation 4.11 *Given an algebra, $\mathcal{A} \subset 2^{\Omega}$, let $\mathbb{S}(\mathcal{A})$ denote the collection of simple random variables from Ω to $\mathbb{C}$. For example if $A \in \mathcal{A}$, then $1_A \in \mathbb{S}(\mathcal{A})$ is a measurable simple function.*

Lemma 4.12. *Let $\mathcal{A} \subset 2^{\Omega}$ be an algebra, then;*

1. *$\mathbb{S}(\mathcal{A})$ is a sub-algebra of all functions from Ω to $\mathbb{C}$.*
2. *$f : \Omega \to \mathbb{C}$, is a $\mathcal{A}$ – simple random variable iff there exists $\alpha_i \in \mathbb{C}$ and $A_i \in \mathcal{A}$ for $1 \le i \le n$ for some $n \in \mathbb{N}$ such that*
$$f = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{4.10}$$
3. *For any function, $F : \mathbb{C} \to \mathbb{C}$, $F \circ f \in \mathbb{S}(\mathcal{A})$ for all $f \in \mathbb{S}(\mathcal{A})$. In particular, $|f| \in \mathbb{S}(\mathcal{A})$ if $f \in \mathbb{S}(\mathcal{A})$.*

Proof. 1. Let us observe that $1_{\Omega} = 1$ and $1_{\emptyset} = 0$ are in $\mathbb{S}(\mathcal{A})$. If $f, g \in \mathbb{S}(\mathcal{A})$ and $c \in \mathbb{C} \setminus \{0\}$, then

$$\{f + cg = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a+cb=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{4.11}$$

and

$$\{f \cdot g = \lambda\} = \bigcup_{a,b \in \mathbb{C}: a \cdot b=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{4.12}$$

from which it follows that $f + cg$ and $f \cdot g$ are back in $\mathbb{S}(\mathcal{A})$.

2. Since $\mathbb{S}(\mathcal{A})$ is an algebra, every f of the form in Eq. (4.10) is in $\mathbb{S}(\mathcal{A})$. Conversely if $f \in \mathbb{S}(\mathcal{A})$ it follows by definition that $f = \sum_{\alpha \in f(\Omega)} \alpha 1_{\{f=\alpha\}}$ which is of the form in Eq. (4.10).

3. If $F : \mathbb{C} \to \mathbb{C}$, then

$$F \circ f = \sum_{\alpha \in f(\Omega)} F(\alpha) \cdot 1_{\{f=\alpha\}} \in \mathbb{S}(\mathcal{A}).$$

■

Exercise 4.1 ($\mathcal{A}$ – measurable simple functions). As in Example 3.19, let $\mathcal{A} \subset 2^X$ be a finite algebra and $\{B_1, \dots, B_k\}$ be the partition of X associated to $\mathcal{A}$. Show that a function, $f : X \to \mathbb{C}$, is an $\mathcal{A}$ – simple function iff f is constant on B_i for each i. Thus any $\mathcal{A}$ – simple function is of the form,

$$f = \sum_{i=1}^{k} \alpha_i 1_{B_i} \tag{4.13}$$

for some $\alpha_i \in \mathbb{C}$.

Corollary 4.13. *Suppose that Λ is a finite set and $Z : X \to \Lambda$ is a function. Let*
$$\mathcal{A} := \mathcal{A}(Z) := Z^{-1}\left(2^{\Lambda}\right) := \left\{Z^{-1}(E) : E \subset \Lambda\right\}.$$
Then $\mathcal{A}$ is an algebra and $f : X \to \mathbb{C}$ is an $\mathcal{A}$ – simple function iff $f = F \circ Z$ for some function $F : \Lambda \to \mathbb{C}$.

Proof. For $\lambda \in \Lambda$, let

$$A_{\lambda} := \{Z = \lambda\} = \{x \in X : Z(x) = \lambda\}.$$

The $\{A_{\lambda}\}_{\lambda \in \Lambda}$ is the partition of X determined by $\mathcal{A}$. Therefore f is an $\mathcal{A}$ – simple function iff $f|_{A_{\lambda}}$ is constant for each $\lambda \in \Lambda$. Let us denote this constant value by $F(\lambda)$. As $Z = \lambda$ on A_{λ}, $F : \Lambda \to \mathbb{C}$ is a function such that $f = F \circ Z$.

Conversely if $F : \Lambda \to \mathbb{C}$ is a function and $f = F \circ Z$, then $f = F(\lambda)$ on A_{λ}, i.e. f is an $\mathcal{A}$ – simple function. ■

4.2.1 The algebraic structure of simple functions*

Definition 4.14. *A* ***simple function algebra***, *$\mathbb{S}$, is a subalgebra*[1] *of the bounded complex functions on X such that $1 \in \mathbb{S}$ and each function in $\mathbb{S}$ is a simple function. If $\mathbb{S}$ is a simple function algebra, let*

$$\mathcal{A}(\mathbb{S}) := \{A \subset X : 1_A \in \mathbb{S}\}.$$

(It is easily checked that $\mathcal{A}(\mathbb{S})$ is a sub-algebra of 2^X.)

[1] To be more explicit we are assuming that $\mathbb{S}$ is a linear subspace of bounded functions which is closed under pointwise multiplication.

Lemma 4.15. *Suppose that* $\mathbb{S}$ *is a simple function algebra,* $f \in \mathbb{S}$ *and* $\alpha \in f(X)$ *– the range of* f. *Then* $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$.

Proof. Let $\{\lambda_i\}_{i=0}^n$ be an enumeration of $f(X)$ with $\lambda_0 = \alpha$. Then

$$g := \left[\prod_{i=1}^n (\alpha - \lambda_i)\right]^{-1} \prod_{i=1}^n (f - \lambda_i 1) \in \mathbb{S}.$$

Moreover, we see that $g = 0$ on $\cup_{i=1}^n \{f = \lambda_i\}$ while $g = 1$ on $\{f = \alpha\}$. So we have shown $g = 1_{\{f=\alpha\}} \in \mathbb{S}$ and therefore that $\{f = \alpha\} \in \mathcal{A}(\mathbb{S})$. ■

Exercise 4.2. Continuing the notation introduced above:

1. Show $\mathcal{A}(\mathbb{S})$ is an algebra of sets.
2. Show $\mathbb{S}(\mathcal{A})$ is a simple function algebra.
3. Show that the map

$$\mathcal{A} \in \left\{\text{Algebras} \subset 2^X\right\} \to \mathbb{S}(\mathcal{A}) \in \{\textbf{simple function algebras on } X\}$$

is bijective and the map, $\mathbb{S} \to \mathcal{A}(\mathbb{S})$, is the inverse map.

Solution to Exercise (4.2).

1. Since $0 = 1_\emptyset, 1 = 1_X \in \mathbb{S}$, it follows that $\emptyset$ and X are in $\mathcal{A}(\mathbb{S})$. If $A \in \mathcal{A}(\mathbb{S})$, then $1_{A^c} = 1 - 1_A \in \mathbb{S}$ and so $A^c \in \mathcal{A}(\mathbb{S})$. Finally, if $A, B \in \mathcal{A}(\mathbb{S})$ then $1_{A\cap B} = 1_A \cdot 1_B \in \mathbb{S}$ and thus $A \cap B \in \mathcal{A}(\mathbb{S})$.
2. If $f, g \in \mathbb{S}(\mathcal{A})$ and $c \in \mathbb{F}$, then

$$\{f + cg = \lambda\} = \bigcup_{a,b\in\mathbb{F}: a+cb=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A}$$

and

$$\{f \cdot g = \lambda\} = \bigcup_{a,b\in\mathbb{F}: a\cdot b=\lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A}$$

from which it follows that $f + cg$ and $f \cdot g$ are back in $\mathbb{S}(\mathcal{A})$.
3. If $f : \Omega \to \mathbb{C}$ is a simple function such that $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$, then $f = \sum_{\lambda\in\mathbb{C}} \lambda 1_{\{f=\lambda\}} \in \mathbb{S}$. Conversely, by Lemma 4.15, if $f \in \mathbb{S}$ then $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$. Therefore, a simple function, $f : X \to \mathbb{C}$ is in $\mathbb{S}$ iff $1_{\{f=\lambda\}} \in \mathbb{S}$ for all $\lambda \in \mathbb{C}$. With this preparation, we are now ready to complete the verification.
First off,

$$A \in \mathcal{A}(\mathbb{S}(\mathcal{A})) \iff 1_A \in \mathbb{S}(\mathcal{A}) \iff A \in \mathcal{A}$$

which shows that $\mathcal{A}(\mathbb{S}(\mathcal{A})) = \mathcal{A}$. Similarly,

$$\begin{aligned} f \in \mathbb{S}(\mathcal{A}(\mathbb{S})) &\iff \{f = \lambda\} \in \mathcal{A}(\mathbb{S}) \ \forall \lambda \in \mathbb{C} \\ &\iff 1_{\{f=\lambda\}} \in \mathbb{S} \ \forall \lambda \in \mathbb{C} \\ &\iff f \in \mathbb{S} \end{aligned}$$

which shows $\mathbb{S}(\mathcal{A}(\mathbb{S})) = \mathbb{S}$.

4.3 Simple Integration

Definition 4.16 (Simple Integral). *Suppose now that* P *is a finitely additive probability measure on an algebra* $\mathcal{A} \subset 2^X$. *For* $f \in \mathbb{S}(\mathcal{A})$ *the* ***integral or expectation,*** $\mathbb{E}(f) = \mathbb{E}_P(f)$, *is defined by*

$$\mathbb{E}_P(f) = \int_X f dP = \sum_{y\in\mathbb{C}} yP(f = y). \tag{4.14}$$

Example 4.17. Suppose that $A \in \mathcal{A}$, then

$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A). \tag{4.15}$$

Remark 4.18. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (2.1) by

$$P(A) = \lim_{N\to\infty} \frac{1}{N} \sum 1_A(\omega(k))$$

where $\omega(k) \in \Omega$ was the result of the k^{th} "independent" experiment. If we use this interpretation back in Eq. (4.14) we arrive at,

$$\begin{aligned} \mathbb{E}(f) = \sum_{y\in\mathbb{C}} yP(f = y) &= \sum_{y\in\mathbb{C}} y \cdot \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{y\in\mathbb{C}} y \sum_{k=1}^N 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \sum_{y\in\mathbb{C}} f(\omega(k)) \cdot 1_{f(\omega(k))=y} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N f(\omega(k)). \end{aligned}$$

Thus informally, $\mathbb{E}f$ should represent the limiting average of the values of f over many "independent" experiments. We will come back to this later when we study the strong law of large numbers.

Proposition 4.19. *The expectation operator,* $\mathbb{E} = \mathbb{E}_P : \mathbb{S}(\mathcal{A}) \to \mathbb{C}$, *satisfies:*

1. *If* $f \in \mathbb{S}(\mathcal{A})$ *and* $\lambda \in \mathbb{C}$, *then*
$$\mathbb{E}(\lambda f) = \lambda \mathbb{E}(f). \tag{4.16}$$
2. *If* $f, g \in \mathbb{S}(\mathcal{A})$, *then*
$$\mathbb{E}(f+g) = \mathbb{E}(g) + \mathbb{E}(f). \tag{4.17}$$
Items 1. and 2. say that $\mathbb{E}(\cdot)$ *is a linear functional on* $\mathbb{S}(\mathcal{A})$.
3. *If* $f = \sum_{j=1}^N \lambda_j 1_{A_j}$ *for some* $\lambda_j \in \mathbb{C}$ *and some* $A_j \in \mathbb{C}$, *then*
$$\mathbb{E}(f) = \sum_{j=1}^N \lambda_j P(A_j). \tag{4.18}$$
4. $\mathbb{E}$ *is* ***positive****, i.e.* $\mathbb{E}(f) \geq 0$ *for all* $0 \leq f \in \mathbb{S}(\mathcal{A})$. *More generally, if* $f, g \in \mathbb{S}(\mathcal{A})$ *and* $f \leq g$, *then* $\mathbb{E}(f) \leq \mathbb{E}(g)$.
5. *For all* $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}f| \leq \mathbb{E}|f|. \tag{4.19}$$

Proof.

1. If $\lambda \neq 0$, then
$$\mathbb{E}(\lambda f) = \sum_{y \in \mathbb{C}} y\ P(\lambda f = y) = \sum_{y \in \mathbb{C}} y\ P(f = y/\lambda)$$
$$= \sum_{z \in \mathbb{C}} \lambda z\ P(f = z) = \lambda \mathbb{E}(f).$$
The case $\lambda = 0$ is trivial.
2. Writing $\{f = a, g = b\}$ for $f^{-1}(\{a\}) \cap g^{-1}(\{b\})$, then
$$\mathbb{E}(f+g) = \sum_{z \in \mathbb{C}} z\ P(f+g=z)$$
$$= \sum_{z \in \mathbb{C}} z\ P\left(\sum_{a+b=z} \{f=a,\ g=b\}\right)$$
$$= \sum_{z \in \mathbb{C}} z \sum_{a+b=z} P(\{f=a,\ g=b\})$$
$$= \sum_{z \in \mathbb{C}} \sum_{a+b=z} (a+b) P(\{f=a,\ g=b\})$$
$$= \sum_{a,b} (a+b) P(\{f=a,\ g=b\}).$$
But
$$\sum_{a,b} aP(\{f=a,\ g=b\}) = \sum_a a \sum_b P(\{f=a,\ g=b\})$$
$$= \sum_a aP(\cup_b \{f=a,\ g=b\})$$
$$= \sum_a aP(\{f=a\}) = \mathbb{E}f$$
and similarly,
$$\sum_{a,b} bP(\{f=a,\ g=b\}) = \mathbb{E}g.$$
Equation (4.17) is now a consequence of the last three displayed equations.
3. If $f = \sum_{j=1}^N \lambda_j 1_{A_j}$, then
$$\mathbb{E}f = \mathbb{E}\left[\sum_{j=1}^N \lambda_j 1_{A_j}\right] = \sum_{j=1}^N \lambda_j \mathbb{E}1_{A_j} = \sum_{j=1}^N \lambda_j P(A_j).$$
4. If $f \geq 0$ then
$$\mathbb{E}(f) = \sum_{a \geq 0} aP(f=a) \geq 0$$
and if $f \leq g$, then $g - f \geq 0$ so that
$$\mathbb{E}(g) - \mathbb{E}(f) = \mathbb{E}(g-f) \geq 0.$$
5. By the triangle inequality,
$$|\mathbb{E}f| = \left|\sum_{\lambda \in \mathbb{C}} \lambda P(f=\lambda)\right| \leq \sum_{\lambda \in \mathbb{C}} |\lambda| P(f=\lambda) = \mathbb{E}|f|,$$
wherein the last equality we have used Eq. (4.18) and the fact that $|f| = \sum_{\lambda \in \mathbb{C}} |\lambda| 1_{f=\lambda}$.

■

Remark 4.20. If Ω is a finite set and $\mathcal{A} = 2^\Omega$, then
$$f(\cdot) = \sum_{\omega \in \Omega} f(\omega) 1_{\{\omega\}}$$
and hence
$$\mathbb{E}_P f = \sum_{\omega \in \Omega} f(\omega) P(\{\omega\}).$$

Remark 4.21. All of the results in Proposition 4.19 and Remark 4.20 remain valid when P is replaced by a finite measure, $\mu : \mathcal{A} \to [0,\infty)$, i.e. it is enough to assume $\mu(X) < \infty$.

Exercise 4.3. Let P is a finitely additive probability measure on an algebra $\mathcal{A} \subset 2^X$ and for $A, B \in \mathcal{A}$ let $\rho(A,B) := P(A \Delta B)$ where $A\Delta B = (A \setminus B) \cup (B \setminus A)$. Show;

1. $\rho(A,B) = \mathbb{E}|1_A - 1_B|$ and then use this (or not) to show
2. $\rho(A,C) \le \rho(A,B) + \rho(B,C)$ for all $A,B,C \in \mathcal{A}$.

Remark: it is now easy to see that $\rho : \mathcal{A} \times \mathcal{A} \to [0,1]$ satisfies the axioms of a metric except for the condition that $\rho(A,B) = 0$ does not imply that $A = B$ but only that $A = B$ modulo a set of probability zero.

Remark 4.22 (Chebyshev's Inequality). Suppose that $f \in \mathbb{S}(\mathcal{A})$, $\varepsilon > 0$, and $p > 0$, then

$$1_{|f|\ge\varepsilon} \le \frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon} \le \varepsilon^{-p}|f|^p$$

and therefore, see item 4. of Proposition 4.19,

$$P(\{|f| \ge \varepsilon\}) = \mathbb{E}\left[1_{|f|\ge\varepsilon}\right] \le \mathbb{E}\left[\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}\right] \le \varepsilon^{-p}\mathbb{E}|f|^p. \tag{4.20}$$

Observe that

$$|f|^p = \sum_{\lambda\in\mathbb{C}} |\lambda|^p 1_{\{f=\lambda\}}$$

is a simple random variable and $\{|f| \ge \varepsilon\} = \sum_{|\lambda|\ge\varepsilon} \{f = \lambda\} \in \mathcal{A}$ as well. Therefore, $\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}$ is still a simple random variable.

Lemma 4.23 (Inclusion Exclusion Formula). *If $A_n \in \mathcal{A}$ for $n = 1,2,\dots,M$ such that $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, then*

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le M} \mu(A_{n_1}\cap\dots\cap A_{n_k}). \tag{4.21}$$

Proof. This may be proved inductively from Eq. (4.2). We will give a different and perhaps more illuminating proof here. Let $A := \cup_{n=1}^M A_n$.

Since $A^c = \left(\cup_{n=1}^M A_n\right)^c = \cap_{n=1}^M A_n^c$, we have

$$\begin{aligned}1 - 1_A = 1_{A^c} &= \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M (1 - 1_{A_n}) \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}}\cdots 1_{A_{n_k}} \\ &= 1 + \sum_{k=1}^M (-1)^k \sum_{1\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}\cap\dots\cap A_{n_k}}\end{aligned}$$

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}\cap\dots\cap A_{n_k}}. \tag{4.22}$$

Integrating this identity with respect to μ gives Eq. (4.21). ■

Remark 4.24. The following identity holds even when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$,

$$\begin{aligned}\mu\left(\cup_{n=1}^M A_n\right) + \sum_{k=2 \text{ \& } k \text{ even}}^M \sum_{1\le n_1<n_2<\dots<n_k\le M} \mu(A_{n_1}\cap\dots\cap A_{n_k}) \\ = \sum_{k=1 \text{ \& } k \text{ odd}}^M \sum_{1\le n_1<n_2<\dots<n_k\le M} \mu(A_{n_1}\cap\dots\cap A_{n_k}).\end{aligned} \tag{4.23}$$

This can be proved by moving every term with a negative sign on the right side of Eq. (4.22) to the left side and then integrate the resulting identity. Alternatively, Eq. (4.23) follows directly from Eq. (4.21) if $\mu\left(\cup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (4.23) are infinite.

To better understand Eq. (4.22), consider the case $M = 3$ where,

$$\begin{aligned}1 - 1_A &= (1 - 1_{A_1})(1 - 1_{A_2})(1 - 1_{A_3}) \\ &= 1 - (1_{A_1} + 1_{A_2} + 1_{A_3}) \\ &\quad + 1_{A_1}1_{A_2} + 1_{A_1}1_{A_3} + 1_{A_2}1_{A_3} - 1_{A_1}1_{A_2}1_{A_3}\end{aligned}$$

so that

$$1_{A_1\cup A_2\cup A_3} = 1_{A_1} + 1_{A_2} + 1_{A_3} - (1_{A_1\cap A_2} + 1_{A_1\cap A_3} + 1_{A_2\cap A_3}) + 1_{A_1\cap A_2\cap A_3}$$

Here is an alternate proof of Eq. (4.22). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1\cap\dots\cap A_m$ and $\omega \notin A_{m+1}\cup\dots\cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (4.22) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

$$\begin{aligned}\sum_{k=1}^{M}(-1)^{k+1}&\sum_{1\le n_1<n_2<\dots<n_k\le M}1_{A_{n_1}\cap\dots\cap A_{n_k}}(\omega)\\&=\sum_{k=1}^{m}(-1)^{k+1}\sum_{1\le n_1<n_2<\dots<n_k\le m}1_{A_{n_1}\cap\dots\cap A_{n_k}}(\omega)\\&=\sum_{k=1}^{m}(-1)^{k+1}\binom{m}{k}\\&=1-\sum_{k=0}^{m}(-1)^{k}(1)^{n-k}\binom{m}{k}\\&=1-(1-1)^m=1.\end{aligned}$$

This verifies Eq. (4.22) since $1_{\cup_{n=1}^{M}A_n}(\omega)=1$.

Example 4.25 (Coincidences). Let Ω be the set of permutations (think of card shuffling), $\omega:\{1,2,\dots,n\}\to\{1,2,\dots,n\}$, and define $P(A):=\frac{\#(A)}{n!}$ to be the uniform distribution (Haar measure) on Ω. We wish to compute the probability of the event, B, that a random permutation fixes some index i. To do this, let $A_i:=\{\omega\in\Omega:\omega(i)=i\}$ and observe that $B=\cup_{i=1}^{n}A_i$. So by the Inclusion Exclusion Formula, we have

$$P(B)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\le i_1<i_2<i_3<\dots<i_k\le n}P(A_{i_1}\cap\dots\cap A_{i_k}).$$

Since

$$\begin{aligned}P(A_{i_1}\cap\dots\cap A_{i_k})&=P(\{\omega\in\Omega:\omega(i_1)=i_1,\dots,\omega(i_k)=i_k\})\\&=\frac{(n-k)!}{n!}\end{aligned}$$

and

$$\#\{1\le i_1<i_2<i_3<\dots<i_k\le n\}=\binom{n}{k},$$

we find

$$P(B)=\sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!}=\sum_{k=1}^{n}(-1)^{k+1}\frac{1}{k!}. \tag{4.24}$$

For large n this gives,

$$P(B)=-\sum_{k=1}^{n}\frac{1}{k!}(-1)^k\cong 1-\sum_{k=0}^{\infty}\frac{1}{k!}(-1)^k=1-e^{-1}\cong 0.632.$$

Example 4.26 (Expected number of coincidences). Continue the notation in Example 4.25. We now wish to compute the expected number of fixed points of a random permutation, ω, i.e. how many cards in the shuffled stack have not moved on average. To this end, let

$$X_i=1_{A_i}$$

and observe that

$$N(\omega)=\sum_{i=1}^{n}X_i(\omega)=\sum_{i=1}^{n}1_{\omega(i)=i}=\#\{i:\omega(i)=i\}.$$

denote the number of fixed points of ω. Hence we have

$$\mathbb{E}N=\sum_{i=1}^{n}\mathbb{E}X_i=\sum_{i=1}^{n}P(A_i)=\sum_{i=1}^{n}\frac{(n-1)!}{n!}=1.$$

Let us check the above formulas when $n=3$. In this case we have

ω	$N(\omega)$
1 2 3	3
1 3 2	1
2 1 3	1
2 3 1	0
3 1 2	0
3 2 1	1

and so

$$P(\exists\text{ a fixed point})=\frac{4}{6}=\frac{2}{3}\cong 0.67\cong 0.632$$

while

$$\sum_{k=1}^{3}(-1)^{k+1}\frac{1}{k!}=1-\frac{1}{2}+\frac{1}{6}=\frac{2}{3}$$

and

$$\mathbb{E}N=\frac{1}{6}(3+1+1+0+0+1)=1.$$

The next three problems generalize the results above. The following notation will be used throughout these exercises.

1. $(\Omega,\mathcal{A},P)$ is a finitely additive probability space, so $P(\Omega)=1$,
2. $A_i\in\mathcal{A}$ for $i=1,2,\dots,n$,
3. $N(\omega):=\sum_{i=1}^{n}1_{A_i}(\omega)=\#\{i:\omega\in A_i\}$, and

4. $\{S_k\}_{k=1}^n$ are given by

$$S_k := \sum_{1 \le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}) = \sum_{\Lambda \subset \{1,2,\dots,n\} \ni |\Lambda| = k} P(\cap_{i \in \Lambda} A_i).$$

Exercise 4.4. For $1 \le k \le n$, show;

1. (as functions on Ω) that

$$\binom{N}{k} = \sum_{\Lambda \subset \{1,2,\dots,n\} \ni |\Lambda| = k} 1_{\cap_{i \in \Lambda} A_i}, \tag{4.25}$$

where by definition

$$\binom{m}{k} = \begin{cases} 0 & \text{if } \quad k > m \\ \frac{m!}{k! \cdot (m-k)!} & \text{if } 1 \le k \le m \\ 1 & \text{if } \quad k = 0 \end{cases}. \tag{4.26}$$

2. Conclude from Eq. (4.25) that for all $z \in \mathbb{C}$,

$$(1+z)^N = 1 + \sum_{k=1}^n z^k \sum_{1 \le i_1 < i_2 < \dots < i_k \le n} 1_{A_{i_1} \cap \dots \cap A_{i_k}} \tag{4.27}$$

provided $(1+z)^0 = 1$ even when $z = -1$.

3. Conclude from Eq. (4.25) that $S_k = \mathbb{E}_P \binom{N}{k}$.

Exercise 4.5. Taking expectations of Eq. (4.27) implies,

$$\mathbb{E}\left[(1+z)^N\right] = 1 + \sum_{k=1}^n S_k z^k. \tag{4.28}$$

Show that setting $z = -1$ in Eq. (4.28) gives another proof of the inclusion exclusion formula. **Hint:** use the definition of the expectation to write out $\mathbb{E}\left[(1+z)^N\right]$ explicitly.

Exercise 4.6. Let $1 \le m \le n$. In this problem you are asked to compute the probability that there are exactly m – coincidences. Namely you should show,

$$P(N = m) = \sum_{k=m}^n (-1)^{k-m} \binom{k}{m} S_k = \sum_{k=m}^n (-1)^{k-m} \binom{k}{m} \sum_{1 \le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Hint: differentiate Eq. (4.28) m times with respect to z and then evaluate the result at $z = -1$. In order to do this you will find it useful to derive formulas for;

$$\frac{d^m}{dz^m}|_{z=-1} (1+z)^n \text{ and } \frac{d^m}{dz^m}|_{z=-1} z^k.$$

Example 4.27. Let us again go back to Example 4.26 where we computed,

$$S_k = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Therefore it follows from Exercise 4.6 that

$$P(\exists \text{ exactly m fixed points}) = P(N = m) = \sum_{k=m}^n (-1)^{k-m} \binom{k}{m} \frac{1}{k!} = \frac{1}{m!} \sum_{k=m}^n (-1)^{k-m} \frac{1}{(k-m)!}.$$

So if n is much bigger than m we may conclude that

$$P(\exists \text{ exactly m fixed points}) \cong \frac{1}{m!} e^{-1}.$$

Let us check our results are consistent with Eq. (4.24);

$$\begin{aligned} P(\exists \text{ a fixed point}) &= \sum_{m=1}^n P(N = m) \\ &= \sum_{m=1}^n \sum_{k=m}^n (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{1 \le m \le k \le n} (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^n \sum_{m=1}^k (-1)^{k-m} \binom{k}{m} \frac{1}{k!} \\ &= \sum_{k=1}^n \left[\sum_{m=0}^k (-1)^{k-m} \binom{k}{m} - (-1)^k \right] \frac{1}{k!} \\ &= -\sum_{k=1}^n (-1)^k \frac{1}{k!} \end{aligned}$$

wherein we have used,

$$\sum_{m=0}^{k}(-1)^{k-m}\binom{k}{m}=(1-1)^k=0.$$

4.3.1 Appendix: Bonferroni Inequalities

In this appendix (see Feller Volume 1., p. 106-111 for more) we want to discuss what happens if we truncate the sums in the inclusion exclusion formula of Lemma 4.23. In order to do this we will need the following lemma whose combinatorial meaning was explained to me by Jeff Remmel.

Lemma 4.28. *Let $n\in\mathbb{N}_0$ and $0\le k\le n$, then*

$$\sum_{l=0}^{k}(-1)^l\binom{n}{l}=(-1)^k\binom{n-1}{k}1_{n>0}+1_{n=0}. \tag{4.29}$$

Proof. The case $n=0$ is trivial. We give two proofs for when $n\in\mathbb{N}$.

First proof. Just use induction on k. When $k=0$, Eq. (4.29) holds since $1=1$. The induction step is as follows,

$$\begin{aligned}\sum_{l=0}^{k+1}(-1)^l\binom{n}{l}&=(-1)^k\binom{n-1}{k}+\binom{n}{k+1}\\&=\frac{(-1)^{k+1}}{(k+1)!}\left[n(n-1)\dots(n-k)-(k+1)(n-1)\dots(n-k)\right]\\&=\frac{(-1)^{k+1}}{(k+1)!}\left[(n-1)\dots(n-k)(n-(k+1))\right]=(-1)^{k+1}\binom{n-1}{k+1}.\end{aligned}$$

Second proof. Let $X=\{1,2,\dots,n\}$ and observe that

$$\begin{aligned}m_k:=\sum_{l=0}^{k}(-1)^l\binom{n}{l}&=\sum_{l=0}^{k}(-1)^l\cdot\#\left(\Lambda\in 2^X:\#(\Lambda)=l\right)\\&=\sum_{\Lambda\in 2^X:\ \#(\Lambda)\le k}(-1)^{\#(\Lambda)}\end{aligned} \tag{4.30}$$

Define $T:2^X\to 2^X$ by

$$T(S)=\begin{cases}S\cup\{1\} \text{ if } 1\notin S\\ S\setminus\{1\} \text{ if } 1\in S\end{cases}.$$

Observe that T is a bijection of 2^X such that T takes even cardinality sets to odd cardinality sets and visa versa. Moreover, if we let

$$\Gamma_k:=\left\{\Lambda\in 2^X:\#(\Lambda)\le k \text{ and } 1\in\Lambda \text{ if } \#(\Lambda)=k\right\},$$

then $T(\Gamma_k)=\Gamma_k$ for all $1\le k\le n$. Since

$$\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(T(\Lambda))}=\sum_{\Lambda\in\Gamma_k}-(-1)^{\#(\Lambda)}$$

we see that $\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=0$. Using this observation with Eq. (4.30) implies

$$m_k=\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}+\sum_{\#(\Lambda)=k\ \&\ 1\notin\Lambda}(-1)^{\#(\Lambda)}=0+(-1)^k\binom{n-1}{k}.$$

■

Corollary 4.29 (Bonferroni Inequalitites). *Let $\mu:\mathcal{A}\to[0,\mu(X)]$ be a finitely additive finite measure on $\mathcal{A}\subset 2^X$, $A_n\in\mathcal{A}$ for $n=1,2,\dots,M$, $N:=\sum_{n=1}^{M}1_{A_n}$, and*

$$S_k:=\sum_{1\le i_1<\dots<i_k\le M}\mu\left(A_{i_1}\cap\dots\cap A_{i_k}\right)=\mathbb{E}_\mu\left[\binom{N}{k}\right].$$

Then for $1\le k\le M$,

$$\mu\left(\cup_{n=1}^M A_n\right)=\sum_{l=1}^{k}(-1)^{l+1}S_l+(-1)^k\,\mathbb{E}_\mu\left[\binom{N-1}{k}\right]. \tag{4.31}$$

This leads to the Bonferroni inequalities;

$$\mu\left(\cup_{n=1}^M A_n\right)\le\sum_{l=1}^{k}(-1)^{l+1}S_l \text{ if } k \text{ is odd}$$

and

$$\mu\left(\cup_{n=1}^M A_n\right)\ge\sum_{l=1}^{k}(-1)^{l+1}S_l \text{ if } k \text{ is even.}$$

Proof. By Lemma 4.28,

$$\sum_{l=0}^{k}(-1)^l\binom{N}{l}=(-1)^k\binom{N-1}{k}1_{N>0}+1_{N=0}.$$

Therefore integrating this equation with respect to μ gives,

$$\mu(X)+\sum_{l=1}^{k}(-1)^l S_l=\mu(N=0)+(-1)^k\,\mathbb{E}_\mu\binom{N-1}{k}$$

and therefore,

$$\mu\left(\cup_{n=1}^{M} A_n\right) = \mu(N>0) = \mu(X) - \mu(N=0) = -\sum_{l=1}^{k} (-1)^l S_l + (-1)^k \mathbb{E}_\mu \binom{N-1}{k}.$$

The Bonferroni inequalities are a simple consequence of Eq. (4.31) and the fact that

$$\binom{N-1}{k} \ge 0 \implies \mathbb{E}_\mu \binom{N-1}{k} \ge 0.$$

■

4.3.2 Appendix: Riemann Stieltjes integral

In this subsection, let X be a set, $\mathcal{A} \subset 2^X$ be an algebra of sets, and $P := \mu : \mathcal{A} \to [0,\infty)$ be a finitely additive measure with $\mu(X) < \infty$. As above let

$$\mathbb{E}_\mu f := \int_X f d\mu := \sum_{\lambda \in \mathbb{C}} \lambda \mu(f = \lambda) \ \forall\ f \in \mathbb{S}(\mathcal{A}). \tag{4.32}$$

Notation 4.30 *For any function, $f : X \to \mathbb{C}$ let $\|f\|_u := \sup_{x\in X} |f(x)|$. Further, let $\bar{\mathbb{S}} := \overline{\mathbb{S}(\mathcal{A})}$ denote those functions, $f : X \to \mathbb{C}$ such that there exists $f_n \in \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$.*

Exercise 4.7. Prove the following statements.

1. For all $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_\mu f| \le \mu(X) \|f\|_u. \tag{4.33}$$
2. If $f \in \bar{\mathbb{S}}$ and $f_n \in \mathbb{S} := \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$, show $\lim_{n\to\infty} \mathbb{E}_\mu f_n$ exists. Also show that defining $\mathbb{E}_\mu f := \lim_{n\to\infty} \mathbb{E}_\mu f_n$ is well defined, i.e. you must show that $\lim_{n\to\infty} \mathbb{E}_\mu f_n = \lim_{n\to\infty} \mathbb{E}_\mu g_n$ if $g_n \in \mathbb{S}$ such that $\lim_{n\to\infty} \|f - g_n\|_u = 0$.
3. Show $\mathbb{E}_\mu : \bar{\mathbb{S}} \to \mathbb{C}$ is still linear and still satisfies Eq. (4.33).
4. Show $|f| \in \bar{\mathbb{S}}$ if $f \in \bar{\mathbb{S}}$ and that Eq. (4.19) is still valid, i.e. $|\mathbb{E}_\mu f| \le \mathbb{E}_\mu |f|$ for all $f \in \bar{\mathbb{S}}$.

Let us now specialize the above results to the case where $X = [0,T]$ for some $T < \infty$. Let $\mathcal{S} := \{(a,b] : 0 \le a \le b \le T\} \cup \{0\}$ which is easily seen to be a semi-algebra. The following proposition is fairly straightforward and will be left to the reader.

Proposition 4.31 (Riemann Stieltjes integral). *Let $F : [0,T] \to \mathbb{R}$ be an increasing function, then;*

1. *there exists a unique finitely additive measure, μ_F, on $\mathcal{A} := \mathcal{A}(\mathcal{S})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $0 \le a \le b \le T$ and $\mu_F(\{0\}) = 0$. (In fact one could allow for $\mu_F(\{0\}) = \lambda$ for any $\lambda \ge 0$, but we would then have to write $\mu_{F,\lambda}$ rather than μ_F.)*
2. *Show $C([0,1],\mathbb{C}) \subset \overline{\mathbb{S}(\mathcal{A})}$. More precisely, suppose $\pi := \{0 = t_0 < t_1 < \dots < t_n = T\}$ is a partition of $[0,T]$ and $c = (c_1,\dots,c_n) \in [0,T]^n$ with $t_{i-1} \le c_i \le t_i$ for each i. Then for $f \in C([0,1],\mathbb{C})$, let*
$$f_{\pi,c} := f(0) 1_{\{0\}} + \sum_{i=1}^{n} f(c_i) 1_{(t_{i-1},t_i]}. \tag{4.34}$$
Show that $\|f - f_{\pi,c}\|_u$ is small provided, $|\pi| := \max\{|t_i - t_{i-1}| : i = 1,2,\dots,n\}$ is small.
3. *Using the above results, show*
$$\int_{[0,T]} f d\mu_F = \lim_{|\pi|\to 0} \sum_{i=1}^{n} f(c_i)(F(t_i) - F(t_{i-1}))$$
where the c_i may be chosen arbitrarily subject to the constraint that $t_{i-1} \le c_i \le t_i$.

It is customary to write $\int_0^T f dF$ for $\int_{[0,T]} f d\mu_F$. This integral satisfies the estimates,

$$\left|\int_{[0,T]} f d\mu_F\right| \le \int_{[0,T]} |f| d\mu_F \le \|f\|_u (F(T) - F(0)) \ \forall\ f \in \overline{\mathbb{S}(\mathcal{A})}.$$

When $F(t) = t$,

$$\int_0^T f dF = \int_0^T f(t) dt,$$

is the usual Riemann integral.

Exercise 4.8. Let $a \in (0,T)$, $\lambda > 0$, and

$$G(x) = \lambda \cdot 1_{x \ge a} = \begin{cases} \lambda \text{ if } x \ge a \\ 0 \text{ if } x < a \end{cases}.$$

1. Explicitly compute $\int_{[0,T]} f d\mu_G$ for all $f \in C([0,1],\mathbb{C})$.
2. If $F(x) = x + \lambda \cdot 1_{x\ge a}$ describe $\int_{[0,T]} f d\mu_F$ for all $f \in C([0,1],\mathbb{C})$. **Hint:** if $F(x) = G(x) + H(x)$ where G and H are two increasing functions on $[0,T]$, show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G + \int_{[0,T]} f d\mu_H.$$

Exercise 4.9. Suppose that $F, G : [0,T] \to \mathbb{R}$ are two increasing functions such that $F(0) = G(0)$, $F(T) = G(T)$, and $F(x) \neq G(x)$ for at most countably many points, $x \in (0,T)$. Show

$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G \text{ for all } f \in C([0,1], \mathbb{C}). \tag{4.35}$$

Note well, given $F(0) = G(0)$, $\mu_F = \mu_G$ on $\mathcal{A}$ iff $F = G$.

One of the points of the previous exercise is to show that Eq. (4.35) holds when $G(x) := F(x+)$ – the right continuous version of F. The exercise applies since and increasing function can have at most countably many jumps ,see Remark **??**. So if we only want to integrate continuous functions, we may always assume that $F : [0,T] \to \mathbb{R}$ is right continuous.

4.4 Simple Independence and the Weak Law of Large Numbers

To motivate the exercises in this section, let us imagine that we are following the outcomes of two "independent" experiments with values $\{\alpha_k\}_{k=1}^\infty \subset \Lambda_1$ and $\{\beta_k\}_{k=1}^\infty \subset \Lambda_2$ where Λ_1 and Λ_2 are two finite set of outcomes. Here we are using term independent in an intuitive form to mean that knowing the outcome of one of the experiments gives us no information about outcome of the other.

As an example of independent experiments, suppose that one experiment is the outcome of spinning a roulette wheel and the second is the outcome of rolling a dice. We expect these two experiments will be independent.

As an example of dependent experiments, suppose that dice roller now has two dice – one red and one black. The person rolling dice throws his black or red dice after the roulette ball has stopped and landed on either black or red respectively. If the black and the red dice are weighted differently, we expect that these two experiments are no longer independent.

Lemma 4.32 (Heuristic). *Suppose that $\{\alpha_k\}_{k=1}^\infty \subset \Lambda_1$ and $\{\beta_k\}_{k=1}^\infty \subset \Lambda_2$ are the outcomes of repeatedly running two experiments independent of each other and for $x \in \Lambda_1$ and $y \in \Lambda_2$,*

$$\begin{aligned} p(x,y) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \alpha_k = x \text{ and } \beta_k = y\}, \\ p_1(x) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \alpha_k = x\}, \text{ and} \\ p_2(y) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \beta_k = y\}. \end{aligned} \tag{4.36}$$

Then $p(x,y) = p_1(x) p_2(y)$. In particular this then implies for any $h : \Lambda_1 \times \Lambda_2 \to \mathbb{R}$ we have,

$$\mathbb{E}h = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N h(\alpha_k, \beta_k) = \sum_{(x,y)\in\Lambda_1\times\Lambda_2} h(x,y) p_1(x) p_2(y).$$

Proof. (Heuristic.) Let us imagine running the first experiment repeatedly with the results being recorded as, $\{\alpha_k^\ell\}_{k=1}^\infty$, where $\ell \in \mathbb{N}$ indicates the ℓ^{th} – run of the experiment. Then we have postulated that, independent of ℓ,

$$p(x,y) := \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\alpha_k^\ell = x \text{ and } \beta_k = y\}} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\alpha_k^\ell = x\}} \cdot 1_{\{\beta_k = y\}}$$

So for any $L \in \mathbb{N}$ we must also have,

$$\begin{aligned} p(x,y) &= \frac{1}{L} \sum_{\ell=1}^L p(x,y) = \frac{1}{L} \sum_{\ell=1}^L \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\alpha_k^\ell = x\}} \cdot 1_{\{\beta_k = y\}} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \frac{1}{L} \sum_{\ell=1}^L 1_{\{\alpha_k^\ell = x\}} \cdot 1_{\{\beta_k = y\}}. \end{aligned}$$

Taking the limit of this equation as $L \to \infty$ and interchanging the order of the limits (this is faith based) implies,

$$p(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\beta_k = y\}} \cdot \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^L 1_{\{\alpha_k^\ell = x\}}. \tag{4.37}$$

Since for fixed k, $\{\alpha_k^\ell\}_{\ell=1}^\infty$ is just another run of the first experiment, by our postulate, we conclude that

$$\lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^L 1_{\{\alpha_k^\ell = x\}} = p_1(x) \tag{4.38}$$

independent of the choice of k. Therefore combining Eqs. (4.36), (4.37), and (4.38) implies,

$$p(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N 1_{\{\beta_k = y\}} \cdot p_1(x) = p_2(y) p_1(x).$$

■

To understand this "Lemma" in another but equivalent way, let $X_1 : \Lambda_1 \times \Lambda_2 \to \Lambda_1$ and $X_2 : \Lambda_1 \times \Lambda_2 \to \Lambda_2$ be the projection maps, $X_1(x,y) = x$ and

$X_2(x,y) = y$ respectively. Further suppose that $f : \Lambda_1 \to \mathbb{R}$ and $g : \Lambda_2 \to \mathbb{R}$ are functions, then using the heuristics Lemma 4.32 implies,

$$\begin{aligned}\mathbb{E}[f(X_1)g(X_2)] &= \sum_{(x,y)\in\Lambda_1\times\Lambda_2} f(x)g(y)p_1(x)p_2(y) \\ &= \sum_{x\in\Lambda_1} f(x)p_1(x)\cdot\sum_{y\in\Lambda_2} g(y)p_2(y) = \mathbb{E}f(X_1)\cdot\mathbb{E}g(X_2).\end{aligned}$$

Hopefully these heuristic computations will convince you that the mathematical notion of independence developed below is relevant. In what follows, we will use the obvious generalization of our "results" above to the setting of n – independent experiments. For notational simplicity we will now assume that $\Lambda_1 = \Lambda_2 = \cdots = \Lambda_n = \Lambda.$

Let Λ be a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, and $X_i : \Omega \to \Lambda$ be defined by $X_i(\omega) = \omega_i$ for $\omega \in \Omega$ and $i = 1,2,\ldots,n$. We further suppose $p : \Omega \to [0,1]$ is a function such that

$$\sum_{\omega\in\Omega} p(\omega) = 1$$

and $P : 2^\Omega \to [0,1]$ is the probability measure defined by

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \in 2^\Omega. \tag{4.39}$$

Exercise 4.10 (Simple Independence 1.). Suppose $q_i : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$ for $i = 1,2,\ldots,n$ and now define $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$. Show for any functions, $f_i : \Lambda \to \mathbb{R}$ that

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] = \prod_{i=1}^n \mathbb{E}_{Q_i} f_i$$

where Q_i is the measure on Λ defined by, $Q_i(\gamma) = \sum_{\lambda\in\gamma} q_i(\lambda)$ for all $\gamma \subset \Lambda$.

Exercise 4.11 (Simple Independence 2.). Prove the converse of the previous exercise. Namely, if

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] \tag{4.40}$$

for any functions, $f_i : \Lambda \to \mathbb{R}$, then there exists functions $q_i : \Lambda \to [0,1]$ with $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$, such that $p(\omega) = \prod_{i=1}^n q_i(\omega_i)$.

Definition 4.33 (Independence). *We say simple random variables, $X_1,\ldots,X_n$ with values in Λ on some probability space, $(\Omega,\mathcal{A},P)$ are independent (more precisely P – independent) if Eq. (4.40) holds for all functions, $f_i : \Lambda \to \mathbb{R}$.*

Exercise 4.12 (Simple Independence 3.). Let $X_1,\ldots,X_n : \Omega \to \Lambda$ and $P : 2^\Omega \to [0,1]$ be as described before Exercise 4.10. Show $X_1,\ldots,X_n$ are independent iff

$$P(X_1 \in A_1,\ldots,X_n \in A_n) = P(X_1 \in A_1)\ldots P(X_n \in A_n) \tag{4.41}$$

for all choices of $A_i \subset \Lambda$. Also explain why it is enough to restrict the A_i to single point subsets of Λ.

Exercise 4.13 (A Weak Law of Large Numbers). Suppose that $\Lambda \subset \mathbb{R}$ is a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, $p(\omega) = \prod_{i=1}^n q(\omega_i)$ where $q : \Lambda \to [0,1]$ such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$, and let $P : 2^\Omega \to [0,1]$ be the probability measure defined as in Eq. (4.39). Further let $X_i(\omega) = \omega_i$ for $i = 1,2,\ldots,n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

$$S_n = \frac{1}{n}(X_1 + \cdots + X_n).$$

1. Show, $\xi = \sum_{\lambda\in\Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda\in\Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda\in\Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{4.42}$$
2. Show, $\mathbb{E}S_n = \xi$.
3. Let $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Show
$$\mathbb{E}[(X_i - \xi)(X_j - \xi)] = \delta_{ij}\sigma^2.$$
4. Using $S_n - \xi$ may be expressed as, $\frac{1}{n}\sum_{i=1}^n (X_i - \xi)$, show
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{4.43}$$
5. Conclude using Eq. (4.43) and Remark 4.22 that
$$P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{4.44}$$

   So for large n, S_n is concentrated near $\xi = \mathbb{E}X_i$ with probability approaching 1 for n large. This is a version of the weak law of large numbers.

Definition 4.34 (Covariance). *Let $(\Omega,\mathcal{B},P)$ is a finitely additive probability. The* ***covariance,*** *$\mathrm{Cov}(X,Y)$, of $X,Y \in \mathbb{S}(\mathcal{B})$ is defined by*

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \xi_X)(Y - \xi_Y)] = \mathbb{E}[XY] - \mathbb{E}X\cdot\mathbb{E}Y$$

where $\xi_X := \mathbb{E}X$ and $\xi_Y := \mathbb{E}Y$. The variance of X,

$$\mathrm{Var}(X) := \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$$

We say that X and Y are ***uncorrelated*** *if* $\mathrm{Cov}(X,Y)=0$, *i.e.* $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. *More generally we say* $\{X_k\}_{k=1}^n \subset \mathbb{S}(\mathcal{B})$ *are uncorrelated iff* $\mathrm{Cov}(X_i, X_j) = 0$ *for all* $i \neq j$.

Remark 4.35. 1. Observe that X and Y are independent iff $f(X)$ and $g(Y)$ are uncorrelated for all functions, f and g on the range of X and Y respectively. In particular if X and Y are independent then $\mathrm{Cov}(X,Y)=0$.

2. If you look at your proof of the weak law of large numbers in Exercise 4.13 you will see that it suffices to assume that $\{X_i\}_{i=1}^n$ are uncorrelated rather than the stronger condition of being independent.

Exercise 4.14 (Bernoulli Random Variables). Let $\Lambda = \{0,1\}$, $X : \Lambda \to \mathbb{R}$ be defined by $X(0) = 0$ and $X(1) = 1$, $x \in [0,1]$, and define $Q = x\delta_1 + (1-x)\delta_0$, i.e. $Q(\{0\}) = 1-x$ and $Q(\{1\}) = x$. Verify,

$$\xi(x) := \mathbb{E}_Q X = x \text{ and}$$
$$\sigma^2(x) := \mathbb{E}_Q (X-x)^2 = (1-x)x \leq 1/4.$$

Theorem 4.36 (Weierstrass Approximation Theorem via Bernstein's Polynomials.). *Suppose that* $f \in C([0,1],\mathbb{C})$ *and*

$$p_n(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Then

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |f(x) - p_n(x)| = 0.$$

(See Theorem **??** *for a multi-dimensional generalization of this theorem.)*

Proof. Let $x \in [0,1]$, $\Lambda = \{0,1\}$, $q(0) = 1-x$, $q(1) = x$, $\Omega = \Lambda^n$, and

$$P_x(\{\omega\}) = q(\omega_1)\dots q(\omega_n) = x^{\sum_{i=1}^n \omega_i} \cdot (1-x)^{1-\sum_{i=1}^n \omega_i}.$$

As above, let $S_n = \frac{1}{n}(X_1 + \dots + X_n)$, where $X_i(\omega) = \omega_i$ and observe that

$$P_x\left(S_n = \frac{k}{n}\right) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Therefore, writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$, we have

$$\mathbb{E}_x[f(S_n)] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} = p_n(x).$$

Hence we find

$$\begin{aligned}|p_n(x) - f(x)| &= |\mathbb{E}_x f(S_n) - f(x)| = |\mathbb{E}_x[f(S_n) - f(x)]| \\ &\leq \mathbb{E}_x |f(S_n) - f(x)| \\ &= \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| \geq \varepsilon] \\ &\quad + \mathbb{E}_x[|f(S_n) - f(x)| : |S_n - x| < \varepsilon] \\ &\leq 2M \cdot P_x(|S_n - x| \geq \varepsilon) + \delta(\varepsilon)\end{aligned}$$

where

$$M := \max_{y\in[0,1]} |f(y)| \text{ and}$$
$$\delta(\varepsilon) := \sup\{|f(y) - f(x)| : x,y \in [0,1] \text{ and } |y-x| \leq \varepsilon\}$$

is the modulus of continuity of f. Now by the above exercises,

$$P_x(|S_n - x| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \quad \text{(see Figure 4.1)}$$

and hence we may conclude that

$$\max_{x\in[0,1]} |p_n(x) - f(x)| \leq \frac{M}{2n\varepsilon^2} + \delta(\varepsilon)$$

and therefore, that

$$\limsup_{n\to\infty} \max_{x\in[0,1]} |p_n(x) - f(x)| \leq \delta(\varepsilon).$$

This completes the proof, since by uniform continuity of f, $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$. ■

4.4.1 Product Measures and Fubini's Theorem

In the last part of this section we will extend some of the above ideas to more general "finitely additive measure spaces." A **finitely additive measure space** is a triple, $(X, \mathcal{A}, \mu)$, where X is a set, $\mathcal{A} \subset 2^X$ is an algebra, and $\mu : \mathcal{A} \to [0,\infty]$ is a finitely additive measure. Let $(Y, \mathcal{B}, \nu)$ be another finitely additive measure space.

Definition 4.37. *Let* $\mathcal{A} \odot \mathcal{B}$ *be the smallest sub-algebra of* $2^{X\times Y}$ *containing all sets of the form* $\mathcal{S} := \{A \times B : A \in \mathcal{A} \text{ and } B \in \mathcal{B}\}$. *As we have seen in Exercise 3.10,* $\mathcal{S}$ *is a semi-algebra and therefore* $\mathcal{A} \odot \mathcal{B}$ *consists of subsets,* $C \subset X \times Y$, *which may be written as;*

$$C = \sum_{i=1}^n A_i \times B_i \text{ with } A_i \times B_i \in \mathcal{S}. \tag{4.45}$$

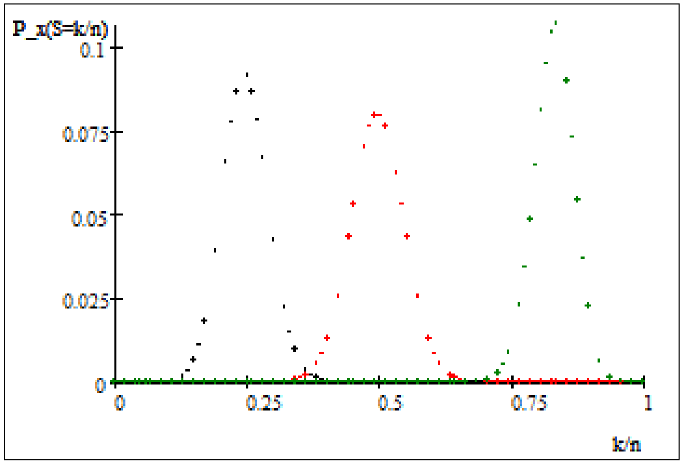


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

Theorem 4.38 (Product Measure and Fubini's Theorem). *Assume that $\mu(X) < \infty$ and $\nu(Y) < \infty$ for simplicity. Then there is a unique finitely additive measure, $\mu \odot \nu$, on $\mathcal{A} \odot \mathcal{B}$ such that $\mu \odot \nu (A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover if $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$ then;*

1. *$y \to f(x, y)$ is in $\mathbb{S}(\mathcal{B})$ for all $x \in X$ and $x \to f(x, y)$ is in $\mathbb{S}(\mathcal{A})$ for all $y \in Y$.*
2. *$x \to \int_Y f(x, y)\, d\nu(y)$ is in $\mathbb{S}(\mathcal{A})$ and $y \to \int_X f(x, y)\, d\mu(x)$ is in $\mathbb{S}(\mathcal{B})$.*
3. *we have,*

$$\int_X \left[\int_Y f(x, y)\, d\nu(y)\right] d\mu(x) = \int_{X\times Y} f(x, y)\, d(\mu \odot \nu)(x, y) = \int_Y \left[\int_X f(x, y)\, d\mu(x)\right] d\nu(y).$$

We will refer to $\mu \odot \nu$ as the ***product measure*** *of μ and ν.*

Proof. According to Eq. (4.45),

$$1_C(x, y) = \sum_{i=1}^n 1_{A_i \times B_i}(x, y) = \sum_{i=1}^n 1_{A_i}(x)\, 1_{B_i}(y)$$

from which it follows that $1_C(x, \cdot) \in \mathbb{S}(\mathcal{B})$ for each $x \in X$ and

$$\int_Y 1_C(x, y)\, d\nu(y) = \sum_{i=1}^n 1_{A_i}(x)\, \nu(B_i).$$

It now follows from this equation that $x \to \int_Y 1_C(x, y)\, d\nu(y) \in \mathbb{S}(\mathcal{A})$ and that

$$\int_X \left[\int_Y 1_C(x, y)\, d\nu(y)\right] d\mu(x) = \sum_{i=1}^n \mu(A_i)\, \nu(B_i).$$

Similarly one shows that

$$\int_Y \left[\int_X 1_C(x, y)\, d\mu(x)\right] d\nu(y) = \sum_{i=1}^n \mu(A_i)\, \nu(B_i).$$

In particular this shows that we may define

$$(\mu \odot \nu)(C) = \sum_{i=1}^n \mu(A_i)\, \nu(B_i)$$

and with this definition we have,

$$\int_X \left[\int_Y 1_C(x, y)\, d\nu(y)\right] d\mu(x) = (\mu \odot \nu)(C) = \int_Y \left[\int_X 1_C(x, y)\, d\mu(x)\right] d\nu(y).$$

From either of these representations it is easily seen that $\mu \odot \nu$ is a finitely additive measure on $\mathcal{A} \odot \mathcal{B}$ with the desired properties. Moreover, we have already verified the Theorem in the special case where $f = 1_C$ with $C \in \mathcal{A} \odot \mathcal{B}$. Since the general element, $f \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$, is a linear combination of such functions, it is easy to verify using the linearity of the integral and the fact that $\mathbb{S}(\mathcal{A})$ and $\mathbb{S}(\mathcal{B})$ are vector spaces that the theorem is true in general. ■

Example 4.39. Suppose that $f \in \mathbb{S}(\mathcal{A})$ and $g \in \mathbb{S}(\mathcal{B})$. Let $f \otimes g(x, y) := f(x)\, g(y)$. Since we have,

$$f \otimes g(x, y) = \left(\sum_a a 1_{f=a}(x)\right)\left(\sum_b b 1_{g=b}(y)\right) = \sum_{a,b} ab 1_{\{f=a\}\times\{g=b\}}(x, y)$$

it follows that $f \otimes g \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$. Moreover, using Fubini's Theorem 4.38 it follows that

$$\int_{X\times Y} f \otimes g\; d(\mu \odot \nu) = \left[\int_X f\; d\mu\right]\left[\int_Y g\; d\nu\right].$$

4.5 Simple Conditional Expectation

In this section, $\mathcal{B}$ is a sub-algebra of 2^{Ω}, $P:\mathcal{B}\to[0,1]$ is a finitely additive probability measure, and $\mathcal{A}\subset\mathcal{B}$ is a finite sub-algebra. As in Example 3.19, for each $\omega\in\Omega$, let $A_\omega := \cap\{A\in\mathcal{A}:\omega\in A\}$ and recall that either $A_\omega = A_{\omega'}$ or $A_\omega\cap A_{\omega'}=\emptyset$ for all $\omega,\omega'\in\Omega$. In particular there is a partition, $\{B_1,\dots,B_n\}$, of Ω such that $A_\omega\in\{B_1,\dots,B_n\}$ for all $\omega\in\Omega$.

Definition 4.40 (Conditional expectation). *Let $X:\Omega\to\mathbb{R}$ be a $\mathcal{B}$ – simple random variable, i.e. $X\in\mathbb{S}(\mathcal{B})$, and*

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)}\mathbb{E}[1_{A_\omega}X] \text{ for all } \omega\in\Omega, \tag{4.46}$$

where by convention, $\bar{X}(\omega)=0$ if $P(A_\omega)=0$. We will denote $\bar{X}$ by $\mathbb{E}[X|\mathcal{A}]$ for $\mathbb{E}_{\mathcal{A}}X$ and call it the conditional expectation of X given $\mathcal{A}$. Alternatively we may write $\bar{X}$ as

$$\bar{X} = \sum_{i=1}^{n}\frac{\mathbb{E}[1_{B_i}X]}{P(B_i)}1_{B_i}, \tag{4.47}$$

again with the convention that $\mathbb{E}[1_{B_i}X]/P(B_i)=0$ if $P(B_i)=0$.

It should be noted, from Exercise 4.1, that $\bar{X}=\mathbb{E}_{\mathcal{A}}X\in\mathbb{S}(\mathcal{A})$. Heuristically, if $(\omega(1),\omega(2),\omega(3),\dots)$ is the sequence of outcomes of "independently" running our "experiment" repeatedly, then

$$\begin{aligned}\bar{X}|_{B_i} &= \frac{\mathbb{E}[1_{B_i}X]}{P(B_i)} \text{ “}=\text{” } \frac{\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{B_i}(\omega(n))X(\omega(n))}{\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{B_i}(\omega(n))} \\ &= \lim_{N\to\infty}\frac{\sum_{n=1}^{N}1_{B_i}(\omega(n))X(\omega(n))}{\sum_{n=1}^{N}1_{B_i}(\omega(n))}.\end{aligned}$$

So to compute $\bar{X}|_{B_i}$ "empirically," we remove all experimental outcomes from the list, $(\omega(1),\omega(2),\omega(3),\dots)\in\Omega^{\mathbb{N}}$, which are not in B_i to form a new list, $(\bar{\omega}(1),\bar{\omega}(2),\bar{\omega}(3),\dots)\in B_i^{\mathbb{N}}$. We then compute $\bar{X}|_{B_i}$ using the empirical formula for the expectation of X relative to the "bar" list, i.e.

$$\bar{X}|_{B_i} = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}X(\bar{\omega}(n)).$$

Exercise 4.15 (Simple conditional expectation). Let $X\in\mathbb{S}(\mathcal{B})$ and, for simplicity, assume all functions are real valued. Prove the following assertions;

1. **(Orthogonal Projection Property 1.)** If $Z\in\mathbb{S}(\mathcal{A})$, then
$$\mathbb{E}[XZ] = \mathbb{E}[\bar{X}Z] = \mathbb{E}[\mathbb{E}_{\mathcal{A}}X\cdot Z] \tag{4.48}$$
and
$$(\mathbb{E}_{\mathcal{A}}Z)(\omega) = \begin{cases} Z(\omega) & \text{if } P(A_\omega)>0 \\ 0 & \text{if } P(A_\omega)=0\end{cases}. \tag{4.49}$$
In particular, $\mathbb{E}_{\mathcal{A}}[\mathbb{E}_{\mathcal{A}}Z]=\mathbb{E}_{\mathcal{A}}Z$.
This basically says that $\mathbb{E}_{\mathcal{A}}$ is orthogonal projection from $\mathbb{S}(\mathcal{B})$ onto $\mathbb{S}(\mathcal{A})$ relative to the inner product
$$(f,g) = \mathbb{E}[fg] \text{ for all } f,g\in\mathbb{S}(\mathcal{B}).$$
2. **(Orthogonal Projection Property 2.)** If $Y\in\mathbb{S}(\mathcal{A})$ satisfies, $\mathbb{E}[XZ]=\mathbb{E}[YZ]$ for all $Z\in\mathbb{S}(\mathcal{A})$, then $Y(\omega)=\bar{X}(\omega)$ whenever $P(A_\omega)>0$. In particular, $P(Y\neq\bar{X})=0$. **Hint:** use item 1. to compute $\mathbb{E}\left[(\bar{X}-Y)^2\right]$.
3. **(Best Approximation Property.)** For any $Y\in\mathbb{S}(\mathcal{A})$,
$$\mathbb{E}\left[(X-\bar{X})^2\right] \le \mathbb{E}\left[(X-Y)^2\right] \tag{4.50}$$
with equality iff $\bar{X}=Y$ almost surely (a.s. for short), where $\bar{X}=Y$ a.s. iff $P(\bar{X}\neq Y)=0$. In words, $\bar{X}=\mathbb{E}_{\mathcal{A}}X$ is the best ("L^2") approximation to X by an $\mathcal{A}$ – measurable random variable.
4. **(Contraction Property.)** $\mathbb{E}|\bar{X}|\le\mathbb{E}|X|$. (It is typically **not** true that $|\bar{X}(\omega)|\le|X(\omega)|$ for all ω.)
5. (**Pull Out Property.**) If $Z\in\mathbb{S}(\mathcal{A})$, then
$$\mathbb{E}_{\mathcal{A}}[ZX] = Z\mathbb{E}_{\mathcal{A}}X.$$

Example 4.41 (Heuristics of independence and conditional expectations). Let us suppose that we have an experiment consisting of spinning a spinner with values in $\Lambda_1=\{1,2,\dots,10\}$ and rolling a die with values in $\Lambda_2=\{1,2,3,4,5,6\}$. So the outcome of an experiment is represented by a point, $\omega=(x,y)\in\Omega=\Lambda_1\times\Lambda_2$. Let $X(x,y)=x$, $Y(x,y)=y$, $\mathcal{B}=2^{\Omega}$, and

$$\mathcal{A} = X^{-1}(2^{\Lambda_1}) = \left\{X^{-1}(A): A\subset\Lambda_1\right\}\subset\mathcal{B},$$

so that $\mathcal{A}$ is the smallest algebra of subsets of Ω such that $\{X=x\}\in\mathcal{A}$ for all $x\in\Lambda_1$. Notice that the partition associated to $\mathcal{A}$ is precisely

$$\{\{X=1\},\{X=2\},\dots,\{X=10\}\}.$$

Let us now suppose that the spins of the spinner are "empirically independent" of the throws of the dice. As usual let us run the experiment repeatedly to produce a sequence of results, $\omega_n = (x_n, y_n)$ for all $n \in \mathbb{N}$. If $g : \Lambda_2 \to \mathbb{R}$ is a function, we have (heuristically) that

$$\begin{aligned}\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) &= \lim_{N\to\infty} \frac{\sum_{n=1}^N g(Y(\omega(n)))\, 1_{X(\omega(n))=x}}{\sum_{n=1}^N 1_{X(\omega(n))=x}} \\ &= \lim_{N\to\infty} \frac{\sum_{n=1}^N g(y_n)\, 1_{x_n=x}}{\sum_{n=1}^N 1_{x_n=x}}.\end{aligned}$$

As $\{y_n\}$ sequence of results is independent of the $\{x_n\}$ we should expect by the usual mantra (i.e. it does not matter which sequence of independent experiments are used to compute the time averages) that

$$\lim_{N\to\infty} \frac{\sum_{n=1}^N g(y_n)\, 1_{x_n=x}}{\sum_{n=1}^N 1_{x_n=x}} = \lim_{N\to\infty} \frac{1}{M(N)} \sum_{n=1}^{M(N)} g(\bar{y}_n) = \mathbb{E}[g(Y)],$$

where $M(N) = \sum_{n=1}^N 1_{x_n=x}$ and $(\bar{y}_1, \bar{y}_2, \dots) = \{y_l : 1_{x_l=x}\}$. (We are also assuming here that $P(X=x) > 0$ so that we expect, $M(N) \sim P(X=x)N$ for N large, in particular $M(N) \to \infty$.) Thus under the assumption that X and Y are describing "independent" experiments we have heuristically deduced that $\mathbb{E}_{\mathcal{A}}[g(Y)] : \Omega \to \mathbb{R}$ is the constant function;

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \mathbb{E}[g(Y)] \text{ for all } (x,y) \in \Omega. \tag{4.51}$$

Let us further observe that if $f : \Lambda_1 \to \mathbb{R}$ is any other function, then $f(X)$ is an $\mathcal{A}$ – simple function and therefore by Eq. (4.51) and Exercise 4.15

$$\mathbb{E}[f(X)]\cdot\mathbb{E}[g(Y)] = \mathbb{E}[f(X)\cdot\mathbb{E}[g(Y)]] = \mathbb{E}[f(X)\cdot\mathbb{E}_{\mathcal{A}}[g(Y)]] = \mathbb{E}[f(X)\cdot g(Y)].$$

This observation along with Exercise 4.12 gives another "proof" of Lemma 4.32.

Lemma 4.42 (Conditional Expectation and Independence). *Let $\Omega = \Lambda_1 \times \Lambda_2$, X, Y, $\mathcal{B} = 2^{\Omega}$, and $\mathcal{A} = X^{-1}(2^{\Lambda_1})$, be as in Example 4.41 above. Assume that $P : \mathcal{B} \to [0,1]$ is a probability measure. If X and Y are P – independent, then Eq. (4.51) holds.*

Proof. From the definitions of conditional expectation and of independence we have,

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \frac{\mathbb{E}[1_{X=x}\cdot g(Y)]}{P(X=x)} = \frac{\mathbb{E}[1_{X=x}]\cdot\mathbb{E}[g(Y)]}{P(X=x)} = \mathbb{E}[g(Y)].$$

■

The following theorem summarizes much of what we (i.e. you) have shown regarding the underlying notion of independence of a pair of simple functions.

Theorem 4.43 (Independence result summary). *Let $(\Omega, \mathcal{B}, P)$ be a finitely additive probability space, Λ be a finite set, and $X, Y : \Omega \to \Lambda$ be two $\mathcal{B}$ – measurable simple functions, i.e. $\{X = x\} \in \mathcal{B}$ and $\{Y = y\} \in \mathcal{B}$ for all $x, y \in \Lambda$. Further let $\mathcal{A} = \mathcal{A}(X) := \mathcal{A}(\{X = x\} : x \in \Lambda)$. Then the following are equivalent;*

1. *$P(X=x, Y=y) = P(X=x)\cdot P(Y=y)$ for all $x \in \Lambda$ and $y \in \Lambda$,*
2. *$\mathbb{E}[f(X)\,g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)]$ for all functions, $f : \Lambda \to \mathbb{R}$ and $g : \Lambda \to \mathbb{R}$,*
3. *$\mathbb{E}_{\mathcal{A}(X)}[g(Y)] = \mathbb{E}[g(Y)]$ for all $g : \Lambda \to \mathbb{R}$, and*
4. *$\mathbb{E}_{\mathcal{A}(Y)}[f(X)] = \mathbb{E}[f(X)]$ for all $f : \Lambda \to \mathbb{R}$.*

We say that X and Y are P – independent if any one (and hence all) of the above conditions holds.

Countably Additive Measures

Let $\mathcal{A} \subset 2^\Omega$ be an algebra and $\mu : \mathcal{A} \to [0,\infty]$ be a finitely additive measure. Recall that μ is a **premeasure** on $\mathcal{A}$ if μ is σ – additive on $\mathcal{A}$. If μ is a premeasure on $\mathcal{A}$ and $\mathcal{A}$ is a σ – algebra (Definition 3.12), we say that μ is a **measure** on $(\Omega, \mathcal{A})$ and that $(\Omega, \mathcal{A})$ is a **measurable space.**

Definition 5.1. *Let $(\Omega, \mathcal{B})$ be a measurable space. We say that $P : \mathcal{B} \to [0,1]$ is a* ***probability measure on*** *$(\Omega, \mathcal{B})$ if P is a measure on $\mathcal{B}$ such that $P(\Omega) = 1$. In this case we say that $(\Omega, \mathcal{B}, P)$ a probability space.*

5.1 Overview

The goal of this chapter is develop methods for proving the existence of desirable probability measures. with the properties that we desire. The main results of this chapter may are summarized in the following theorem.

Theorem 5.2. *The finitely additive probability measure P on $\mathcal{A}$ extends to σ – additive measure on $\sigma(\mathcal{A})$ iff P is a premeasure on $\mathcal{A}$. If the extension exists it is unique.*

Proof. The uniqueness assertion is proved Proposition 5.15 below. The existence assertion of the theorem in the content of Theorem 5.27. ■

In order to use this theorem it is necessary to determine when a finitely additive probability measure in is in fact a premeasure. The following Proposition is sometimes useful in this regard.

Proposition 5.3 (Equivalent premeasure conditions). *Suppose that P is a finitely additive probability measure on an algebra, $\mathcal{A} \subset 2^\Omega$. Then the following are equivalent:*

1. *P is a premeasure on $\mathcal{A}$, i.e. P is σ – additive on $\mathcal{A}$.*
2. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, $P(A_n) \uparrow P(A)$.*
3. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, $P(A_n) \downarrow P(A)$.*
4. *For all $A_n \in \mathcal{A}$ such that $A_n \uparrow \Omega$, $P(A_n) \uparrow 1$.*
5. *For all $A_n \in \mathcal{A}$ such that $A_n \downarrow \emptyset$, $P(A_n) \downarrow 0$.*

Proof. We will start by showing $1 \iff 2 \iff 3$.

$1. \implies 2.$ Suppose $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Let $A'_n := A_n \setminus A_{n-1}$ with $A_0 := \emptyset$. Then $\{A'_n\}_{n=1}^\infty$ are disjoint, $A_n = \cup_{k=1}^n A'_k$ and $A = \cup_{k=1}^\infty A'_k$. Therefore,

$$P(A) = \sum_{k=1}^\infty P(A'_k) = \lim_{n\to\infty} \sum_{k=1}^n P(A'_k) = \lim_{n\to\infty} P\left(\cup_{k=1}^n A'_k\right) = \lim_{n\to\infty} P(A_n).$$

$2. \implies 1.$ If $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ are disjoint and $A := \cup_{n=1}^\infty A_n \in \mathcal{A}$, then $\cup_{n=1}^N A_n \uparrow A$. Therefore,

$$P(A) = \lim_{N\to\infty} P\left(\cup_{n=1}^N A_n\right) = \lim_{N\to\infty} \sum_{n=1}^N P(A_n) = \sum_{n=1}^\infty P(A_n).$$

$2. \implies 3.$ If $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, then $A_n^c \uparrow A^c$ and therefore,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

$3. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A_n^c \downarrow A^c$ and therefore we again have,

$$\lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n^c) = P(A^c) = 1 - P(A).$$

The same proof used for $2. \iff 3.$ shows $4. \iff 5$ and it is clear that $3. \implies 5.$ To finish the proof we will show $5. \implies 2.$

$5. \implies 2.$ If $A_n \in \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$, then $A \setminus A_n \downarrow \emptyset$ and therefore

$$\lim_{n\to\infty} [P(A) - P(A_n)] = \lim_{n\to\infty} P(A \setminus A_n) = 0.$$

■

Remark 5.4. Observe that the equivalence of items 1. and 2. in the above proposition hold without the restriction that $P(\Omega) = 1$ and in fact $P(\Omega) = \infty$ may be allowed for this equivalence.

Lemma 5.5. *If $\mu : \mathcal{A} \to [0,\infty]$ is a premeasure, then μ is countably sub-additive on $\mathcal{A}$.*

Proof. Suppose that $A_n \in \mathcal{A}$ with $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$. Let $A_1^{'} := A_1$ and for $n \geq 2$, let $A_n^{\prime} := A_n \setminus (A_1 \cup \ldots A_{n-1}) \in \mathcal{A}$. Then $\cup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n^{\prime}$ and therefore by the countable additivity and monotonicity of μ we have,

$$\mu\left(\cup_{n=1}^{\infty} A_n\right) = \mu\left(\sum_{n=1}^{\infty} A_n^{\prime}\right) = \sum_{n=1}^{\infty} \mu\left(A_n^{\prime}\right) \leq \sum_{n=1}^{\infty} \mu\left(A_n\right).$$

■

Let us now specialize to the case where $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{A}\left(\{(a,b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}\right)$. In this case we will describe probability measures, P, on $\mathcal{B}_{\mathbb{R}}$ by their "cumulative distribution functions."

Definition 5.6. *Given a probability measure, P on $\mathcal{B}_{\mathbb{R}}$, the* ***cumulative distribution function*** *(CDF) of P is defined as the function, $F = F_P : \mathbb{R} \to [0,1]$ given as*

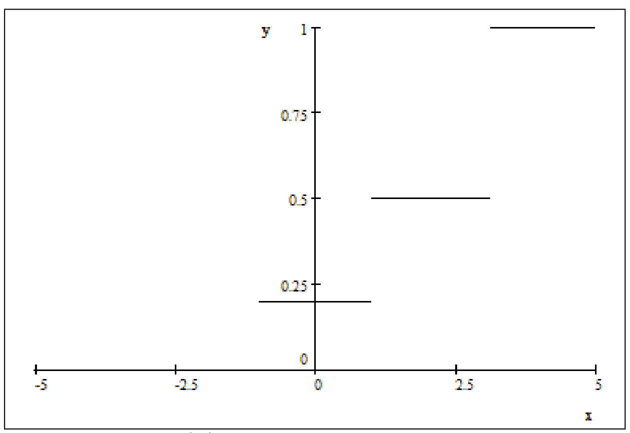
$$F(x) := P((-\infty, x]). \tag{5.1}$$

Example 5.7. Suppose that

$$P = p\delta_{-1} + q\delta_1 + r\delta_\pi$$

with $p, q, r > 0$ and $p + q + r = 1$. In this case,

$$F(x) = \begin{cases} 0 & \text{for } \quad x < -1 \\ p & \text{for } -1 \leq x < 1 \\ p+q & \text{for } \quad 1 \leq x < \pi \\ 1 & \text{for } \pi \leq x < \infty \end{cases}.$$



A plot of $F(x)$ with $p = .2$, $q = .3$, and $r = .5$.

Lemma 5.8. *If $F = F_P : \mathbb{R} \to [0,1]$ is a distribution function for a probability measure, P, on $\mathcal{B}_{\mathbb{R}}$, then:*

1. *F is non-decreasing,*
2. *F is right continuous,*
3. *$F(-\infty) := \lim_{x \to -\infty} F(x) = 0$, and $F(\infty) := \lim_{x \to \infty} F(x) = 1$.*

Proof. The monotonicity of P shows that $F(x)$ in Eq. (5.1) is non-decreasing. For $b \in \mathbb{R}$ let $A_n = (-\infty, b_n]$ with $b_n \downarrow b$ as $n \to \infty$. The continuity of P implies

$$F(b_n) = P((-\infty, b_n]) \downarrow \mu((-\infty, b]) = F(b).$$

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $F(b+) := \lim_{y \downarrow b} F(y) = F(b)$. This show that F is right continuous. Similar arguments show that $F(\infty) = 1$ and $F(-\infty) = 0$. ■

It turns out that Lemma 5.8 has the following important converse.

Theorem 5.9. *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 3.. in Lemma 5.8, there exists a unique probability measure, P_F, on $\mathcal{B}_{\mathbb{R}}$ such that*

$$P_F((a,b]) = F(b) - F(a) \text{ for all } -\infty < a \leq b < \infty.$$

Proof. The uniqueness assertion is proved in Corollary 5.17 below or see Exercises 5.2 and 5.11 below. The existence portion of the theorem is a special case of Theorem 5.33 below. ■

Example 5.10 (Uniform Distribution). The function,

$$F(x) := \begin{cases} 0 & \text{for } \quad x \leq 0 \\ x & \text{for } \ 0 \leq x < 1 \\ 1 & \text{for } \ 1 \leq x < \infty \end{cases},$$

is the distribution function for a measure, m on $\mathcal{B}_{\mathbb{R}}$ which is concentrated on $(0,1]$. The measure, m is called the **uniform distribution** or **Lebesgue measure** on $(0,1]$.

With this summary in hand, let us now start the formal development. We begin with uniqueness statement in Theorem 5.2.

5.2 $\pi - \lambda$ Theorem

Recall that a collection, $\mathcal{P} \subset 2^{\Omega}$, is a π – **class** or π – system if it is closed under finite intersections. We also need the notion of a λ –system.

Definition 5.11 (λ – system). *A collection of sets, $\mathcal{L} \subset 2^{\Omega}$, is λ –* ***class*** *or λ –* ***system*** *if*

a. $\Omega \in \mathcal{L}$

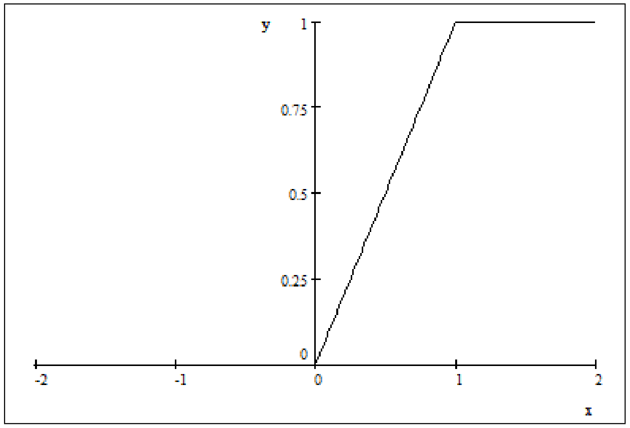


Fig. 5.1. The cumulative distribution function for the uniform distribution.

b. If $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$. (Closed under proper differences.)
c. If $A_n \in \mathcal{L}$ and $A_n \uparrow A$, then $A \in \mathcal{L}$. (Closed under countable increasing unions.)

Remark 5.12. If $\mathcal{L}$ is a collection of subsets of Ω which is both a λ – class and a π – system then $\mathcal{L}$ is a σ – algebra. Indeed, since $A^c = \Omega \setminus A$, we see that any λ - system is closed under complementation. If $\mathcal{L}$ is also a π – system, it is closed under intersections and therefore $\mathcal{L}$ is an algebra. Since $\mathcal{L}$ is also closed under increasing unions, $\mathcal{L}$ is a σ – algebra.

Lemma 5.13 (Alternate Axioms for a λ – System*). *Suppose that $\mathcal{L} \subset 2^\Omega$ is a collection of subsets Ω. Then $\mathcal{L}$ is a λ – class iff λ satisfies the following postulates:*

1. *$\Omega \in \mathcal{L}$*
2. *$A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$. (Closed under complementation.)*
3. *If $\{A_n\}_{n=1}^\infty \subset \mathcal{L}$ are disjoint, then $\sum_{n=1}^\infty A_n \in \mathcal{L}$. (Closed under disjoint unions.)*

Proof. Suppose that $\mathcal{L}$ satisfies a. – c. above. Clearly then postulates 1. and 2. hold. Suppose that $A, B \in \mathcal{L}$ such that $A \cap B = \emptyset$, then $A \subset B^c$ and

$$A^c \cap B^c = B^c \setminus A \in \mathcal{L}.$$

Taking complements of this result shows $A \cup B \in \mathcal{L}$ as well. So by induction, $B_m := \sum_{n=1}^m A_n \in \mathcal{L}$. Since $B_m \uparrow \sum_{n=1}^\infty A_n$ it follows from postulate c. that $\sum_{n=1}^\infty A_n \in \mathcal{L}$.

Now suppose that $\mathcal{L}$ satisfies postulates 1. – 3. above. Notice that $\emptyset \in \mathcal{L}$ and by postulate 3., $\mathcal{L}$ is closed under finite disjoint unions. Therefore if $A, B \in \mathcal{L}$ with $A \subset B$, then $B^c \in \mathcal{L}$ and $A \cap B^c = \emptyset$ allows us to conclude that $A \cup B^c \in \mathcal{L}$. Taking complements of this result shows $B \setminus A = A^c \cap B \in \mathcal{L}$ as well, i.e. postulate b. holds. If $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $B_n := A_n \setminus A_{n-1} \in \mathcal{L}$ for all n, where by convention $A_0 = \emptyset$. Hence it follows by postulate 3 that $\cup_{n=1}^\infty A_n = \sum_{n=1}^\infty B_n \in \mathcal{L}$. ■

Theorem 5.14 (Dynkin's $\pi - \lambda$ Theorem). *If $\mathcal{L}$ is a λ class which contains a contains a π – class, $\mathcal{P}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

Proof. We start by proving the following assertion; for any element $C \in \mathcal{L}$, the collection of sets,

$$\mathcal{L}^C := \{D \in \mathcal{L} : C \cap D \in \mathcal{L}\},$$

is a λ – system. To prove this claim, observe that: a. $\Omega \in \mathcal{L}^C$, b. if $A \subset B$ with $A, B \in \mathcal{L}^C$, then $A \cap C$, $B \cap C \in \mathcal{L}$ with $A \cap C \subset B \cap C$ and therefore,

$$(B \setminus A) \cap C = [B \cap C] \setminus A = [B \cap C] \setminus [A \cap C] \in \mathcal{L}.$$

This shows that $\mathcal{L}^C$ is closed under proper differences. c. If $A_n \in \mathcal{L}^C$ with $A_n \uparrow A$, then $A_n \cap C \in \mathcal{L}$ and $A_n \cap C \uparrow A \cap C \in \mathcal{L}$, i.e. $A \in \mathcal{L}^C$. Hence we have verified $\mathcal{L}^C$ is still a λ – system.

For the rest of the proof, we may assume without loss of generality that $\mathcal{L}$ is the smallest λ – class containing $\mathcal{P}$ – if not just replace $\mathcal{L}$ by the intersection of all λ – classes containing $\mathcal{P}$. Then for $C \in \mathcal{P}$ we know that $\mathcal{L}^C \subset \mathcal{L}$ is a λ - class containing $\mathcal{P}$ and hence $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{P}$ was arbitrary, we have shown, $C \cap D \in \mathcal{L}$ for all $C \in \mathcal{P}$ and $D \in \mathcal{L}$. We may now conclude that if $C \in \mathcal{L}$, then $\mathcal{P} \subset \mathcal{L}^C \subset \mathcal{L}$ and hence again $\mathcal{L}^C = \mathcal{L}$. Since $C \in \mathcal{L}$ is arbitrary, we have shown $C \cap D \in \mathcal{L}$ for all $C, D \in \mathcal{L}$, i.e. $\mathcal{L}$ is a π – system. So by Remark 5.12, $\mathcal{L}$ is a σ algebra. Since $\sigma(\mathcal{P})$ is the smallest σ – algebra containing $\mathcal{P}$ it follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$. ■

As an immediate corollary, we have the following uniqueness result.

Proposition 5.15. *Suppose that $\mathcal{P} \subset 2^\Omega$ is a π – system. If P and Q are two probability[1] measures on $\sigma(\mathcal{P})$ such that $P = Q$ on $\mathcal{P}$, then $P = Q$ on $\sigma(\mathcal{P})$.*

Proof. Let $\mathcal{L} := \{A \in \sigma(\mathcal{P}) : P(A) = Q(A)\}$. One easily shows $\mathcal{L}$ is a λ – class which contains $\mathcal{P}$ by assumption. Indeed, $\Omega \in \mathcal{P} \subset \mathcal{L}$, if $A, B \in \mathcal{L}$ with $A \subset B$, then

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$$

[1] More generally, P and Q could be two measures such that $P(\Omega) = Q(\Omega) < \infty$.

so that $B \setminus A \in \mathcal{L}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $P(A) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} Q(A_n) = Q(A)$ which shows $A \in \mathcal{L}$. Therefore $\sigma(\mathcal{P}) \subset \mathcal{L} = \sigma(\mathcal{P})$ and the proof is complete. ■

Example 5.16. Let $\Omega := \{a, b, c, d\}$ and let μ and ν be the probability measure on 2^Ω determined by, $\mu(\{x\}) = \frac{1}{4}$ for all $x \in \Omega$ and $\nu(\{a\}) = \nu(\{d\}) = \frac{1}{8}$ and $\nu(\{b\}) = \nu(\{c\}) = 3/8$. In this example,

$$\mathcal{L} := \left\{A \in 2^\Omega : P(A) = Q(A)\right\}$$

is λ – system which is not an algebra. Indeed, $A = \{a, b\}$ and $B = \{a, c\}$ are in $\mathcal{L}$ but $A \cap B \notin \mathcal{L}$.

Exercise 5.1. Suppose that μ and ν are two measures (not assumed to be finite) on a measure space, $(\Omega, \mathcal{B})$ such that $\mu = \nu$ on a π – system, $\mathcal{P}$. Further assume $\mathcal{B} = \sigma(\mathcal{P})$ and there exists $\Omega_n \in \mathcal{P}$ such that; i) $\mu(\Omega_n) = \nu(\Omega_n) < \infty$ for all n and ii) $\Omega_n \uparrow \Omega$ as $n \uparrow \infty$. Show $\mu = \nu$ on $\mathcal{B}$.

Hint: Consider the measures, $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$.

Solution to Exercise (5.1). Let $\mu_n(A) := \mu(A \cap \Omega_n)$ and $\nu_n(A) = \nu(A \cap \Omega_n)$ for all $A \in \mathcal{B}$. Then μ_n and ν_n are finite measure such $\mu_n(\Omega) = \nu_n(\Omega)$ and $\mu_n = \nu_n$ on $\mathcal{P}$. Therefore by Proposition 5.15, $\mu_n = \nu_n$ on $\mathcal{B}$. So by the continuity properties of μ and ν, it follows that

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \mu_n(A) = \lim_{n\to\infty} \nu_n(A) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A)$$

for all $A \in \mathcal{B}$.

Corollary 5.17. *A probability measure, P, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is uniquely determined by its cumulative distribution function,*

$$F(x) := P((-\infty, x]).$$

Proof. This follows from Proposition 5.15 wherein we use the fact that $\mathcal{P} := \{(-\infty, x] : x \in \mathbb{R}\}$ is a π – system such that $\mathcal{B}_\mathbb{R} = \sigma(\mathcal{P})$. ■

Remark 5.18. Corollary 5.17 generalizes to $\mathbb{R}^n$. Namely a probability measure, P, on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is uniquely determined by its CDF,

$$F(x) := P((-\infty, x]) \text{ for all } x \in \mathbb{R}^n$$

where now

$$(-\infty, x] := (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n].$$

5.2.1 A Density Result*

Exercise 5.2 (Density of $\mathcal{A}$ in $\sigma(\mathcal{A})$). Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(\mathcal{A})$, and P is a probability measure on $\mathcal{B}$. Let $\rho(A, B) := P(A \Delta B)$. The goal of this exercise is to use the $\pi - \lambda$ theorem to show that $\mathcal{A}$ is dense in $\mathcal{B}$ relative to the "metric," ρ. More precisely you are to show using the following outline that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \triangle B) < \varepsilon$.

1. Recall from Exercise 4.3 that $\rho(a, B) = P(A \triangle B) = \mathbb{E}|1_A - 1_B|$.
2. Observe; if $B = \cup B_i$ and $A = \cup_i A_i$, then
$$B \setminus A = \cup_i [B_i \setminus A] \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and}$$
$$A \setminus B = \cup_i [A_i \setminus B] \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i$$
so that
$$A \triangle B \subset \cup_i (A_i \triangle B_i).$$
3. We also have
$$\begin{aligned}(B_2 \setminus B_1) \setminus (A_2 \setminus A_1) &= B_2 \cap B_1^c \cap (A_2 \setminus A_1)^c \\ &= B_2 \cap B_1^c \cap (A_2 \cap A_1^c)^c \\ &= B_2 \cap B_1^c \cap (A_2^c \cup A_1) \\ &= [B_2 \cap B_1^c \cap A_2^c] \cup [B_2 \cap B_1^c \cap A_1] \\ &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1)\end{aligned}$$
and similarly,
$$(A_2 \setminus A_1) \setminus (B_2 \setminus B_1) \subset (A_2 \setminus B_2) \cup (B_1 \setminus A_1)$$
so that
$$\begin{aligned}(A_2 \setminus A_1) \triangle (B_2 \setminus B_1) &\subset (B_2 \setminus A_2) \cup (A_1 \setminus B_1) \cup (A_2 \setminus B_2) \cup (B_1 \setminus A_1) \\ &= (A_1 \triangle B_1) \cup (A_2 \triangle B_2).\end{aligned}$$
4. Observe that $A_n \in \mathcal{B}$ and $A_n \uparrow A$, then
$$\begin{aligned}P(B \triangle A_n) &= P(B \setminus A_n) + P(A_n \setminus B) \\ &\to P(B \setminus A) + P(A \setminus B) = P(A \triangle B).\end{aligned}$$
5. Let $\mathcal{L}$ be the collection of sets $B \in \mathcal{B}$ for which the assertion of the theorem holds. Show $\mathcal{L}$ is a λ – system which contains $\mathcal{A}$.

Solution to Exercise (5.2). Since $\mathcal{L}$ contains the π – system, $\mathcal{A}$ it suffices by the $\pi - \lambda$ theorem to show $\mathcal{L}$ is a λ – system. Clearly, $\Omega \in \mathcal{L}$ since $\Omega \in \mathcal{A} \subset \mathcal{L}$. If $B_1 \subset B_2$ with $B_i \in \mathcal{L}$ and $\varepsilon > 0$, there exists $A_i \in \mathcal{A}$ such that $P(B_i \triangle A_i) = \mathbb{E}_P |1_{A_i} - 1_{B_i}| < \varepsilon/2$ and therefore,

$$\begin{aligned} P((B_2 \setminus B_1) \triangle (A_2 \setminus A_1)) &\leq P((A_1 \triangle B_1) \cup (A_2 \triangle B_2)) \\ &\leq P((A_1 \triangle B_1)) + P((A_2 \triangle B_2)) < \varepsilon. \end{aligned}$$

Also if $B_n \uparrow B$ with $B_n \in \mathcal{L}$, there exists $A_n \in \mathcal{A}$ such that $P(B_n \triangle A_n) < \varepsilon 2^{-n}$ and therefore,

$$P([\cup_n B_n] \triangle [\cup_n A_n]) \leq \sum_{n=1}^{\infty} P(B_n \triangle A_n) < \varepsilon.$$

Moreover, if we let $B := \cup_n B_n$ and $A^N := \cup_{n=1}^N A_n$, then

$$P\left(B \triangle A^N\right) = P\left(B \setminus A^N\right) + P\left(A^N \setminus B\right) \to P(B \setminus A) + P(A \setminus B) = P(B \triangle A)$$

where $A := \cup_n A_n$. Hence it follows for N large enough that $P\left(B \triangle A^N\right) < \varepsilon$. Since $\varepsilon > 0$ was arbitrary we have shown $B \in \mathcal{L}$ as desired.

5.3 Construction of Measures

Definition 5.19. *Given a collection of subsets, $\mathcal{E}$, of Ω, let $\mathcal{E}_\sigma$ denote the collection of subsets of Ω which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_\delta$ denote the collection of subsets of Ω which are finite or countable intersections of sets from $\mathcal{E}$. We also write $\mathcal{E}_{\sigma\delta} = (\mathcal{E}_\sigma)_\delta$ and $\mathcal{E}_{\delta\sigma} = (\mathcal{E}_\delta)_\sigma$, etc.*

Lemma 5.20. *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra. Then:*

1. *$\mathcal{A}_\sigma$ is closed under taking countable unions and finite intersections.*
2. *$\mathcal{A}_\delta$ is closed under taking countable intersections and finite unions.*
3. *$\{A^c : A \in \mathcal{A}_\sigma\} = \mathcal{A}_\delta$ and $\{A^c : A \in \mathcal{A}_\delta\} = \mathcal{A}_\sigma$.*

Proof. By construction $\mathcal{A}_\sigma$ is closed under countable unions. Moreover if $A = \cup_{i=1}^\infty A_i$ and $B = \cup_{j=1}^\infty B_j$ with $A_i, B_j \in \mathcal{A}$, then

$$A \cap B = \cup_{i,j=1}^\infty A_i \cap B_j \in \mathcal{A}_\sigma,$$

which shows that $\mathcal{A}_\sigma$ is also closed under finite intersections. Item 3. is straight forward and item 2. follows from items 1. and 3. ■

Remark 5.21. Let us recall from Proposition 5.3 and Remark 5.4 that a finitely additive measure $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Furthermore if $\mu(\Omega) < \infty$, then μ is a premeasure on $\mathcal{A}$ iff $\mu(A_n) \downarrow 0$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \downarrow \emptyset$.

Proposition 5.22. *Given a premeasure, $\mu : \mathcal{A} \to [0, \infty]$, we extend μ to $\mathcal{A}_\sigma$ by defining*

$$\mu(B) := \sup\{\mu(A) : \mathcal{A} \ni A \subset B\}. \tag{5.2}$$

This function $\mu : \mathcal{A}_\sigma \to [0, \infty]$ then satisfies;

1. ***(Monotonicity)** If $A, B \in \mathcal{A}_\sigma$ with $A \subset B$ then $\mu(A) \leq \mu(B)$.*
2. ***(Continuity)** If $A_n \in \mathcal{A}$ and $A_n \uparrow A \in \mathcal{A}_\sigma$, then $\mu(A_n) \uparrow \mu(A)$ as $n \to \infty$.*
3. ***(Strong Additivity)** If $A, B \in \mathcal{A}_\sigma$, then*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{5.3}$$
4. ***(Sub-Additivity on $\mathcal{A}_\sigma$)** The function μ is sub-additive on $\mathcal{A}_\sigma$, i.e. if $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$, then*
$$\mu(\cup_{n=1}^\infty A_n) \leq \sum_{n=1}^\infty \mu(A_n). \tag{5.4}$$
5. *(σ - Additivity **on** $\mathcal{A}_\sigma$) The function μ is countably additive on $\mathcal{A}_\sigma$.*

Proof. 1. and 2. Monotonicity follows directly from Eq. (5.2) which then implies $\mu(A_n) \leq \mu(B)$ for all n. Therefore $M := \lim_{n\to\infty} \mu(A_n) \leq \mu(B)$. To prove the reverse inequality, let $\mathcal{A} \ni A \subset B$. Then by the continuity of μ on $\mathcal{A}$ and the fact that $A_n \cap A \uparrow A$ we have $\mu(A_n \cap A) \uparrow \mu(A)$. As $\mu(A_n) \geq \mu(A_n \cap A)$ for all n it then follows that $M := \lim_{n\to\infty} \mu(A_n) \geq \mu(A)$. As $A \in \mathcal{A}$ with $A \subset B$ was arbitrary we may conclude,

$$\mu(B) = \sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \leq M.$$

3. Suppose that $A, B \in \mathcal{A}_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,

$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$

proves Eq. (5.3). In particular, it follows that μ is finitely additive on $\mathcal{A}_\sigma$.

4 and 5. Let $\{A_n\}_{n=1}^\infty$ be any sequence in $\mathcal{A}_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset \mathcal{A}$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,

$$\mu\left(\cup_{n=1}^N A_{n,N}\right) \leq \sum_{n=1}^N \mu(A_{n,N}) \leq \sum_{n=1}^N \mu(A_n) \leq \sum_{n=1}^\infty \mu(A_n). \tag{5.5}$$

Since $\mathcal{A} \ni \cup_{n=1}^N A_{n,N} \uparrow \cup_{n=1}^\infty A_n \in \mathcal{A}_\sigma$, we may let $N \to \infty$ in Eq. (5.5) to conclude Eq. (5.4) holds. If we further assume that $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$ are pairwise disjoint, by the finite additivity and monotonicity of μ on $\mathcal{A}_\sigma$, we have

$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n) = \lim_{N\to\infty} \mu\left(\cup_{n=1}^N A_n\right) \le \mu\left(\cup_{n=1}^\infty A_n\right).$$

This inequality along with Eq. (5.4) shows that μ is σ – additive on $\mathcal{A}_\sigma$. ■

Suppose μ is a **finite** premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and $A \in \mathcal{A}_\delta \cap \mathcal{A}_\sigma$. Since $A, A^c \in \mathcal{A}_\sigma$ and $\Omega = A \cup A^c$, it follows that $\mu(\Omega) = \mu(A) + \mu(A^c)$. From this observation we may extend μ to a function on $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ by defining

$$\mu(A) := \mu(\Omega) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{5.6}$$

Lemma 5.23. *Suppose μ is a finite premeasure on an algebra, $\mathcal{A} \subset 2^\Omega$, and μ has been extended to $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ as described in Proposition 5.22 and Eq. (5.6) above.*

1. *If $A \in \mathcal{A}_\delta$ then $\mu(A) = \inf\{\mu(B) : A \subset B \in \mathcal{A}\}$.*
2. *If $A \in \mathcal{A}_\delta$ and $A_n \in \mathcal{A}$ such that $A_n \downarrow A$, then $\mu(A) =\downarrow \lim_{n\to\infty} \mu(A_n)$.*
3. *μ is strongly additive when restricted to $\mathcal{A}_\delta$.*
4. *If $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset C$, then $\mu(C \setminus A) = \mu(C) - \mu(A)$.*

Proof.

1. Since $\mu(B) = \mu(\Omega) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that

$$\begin{aligned}\inf\{\mu(B) : A \subset B \in \mathcal{A}\} &= \inf\{\mu(\Omega) - \mu(B^c) : \mathcal{A} \ni B^c \subset A^c\} \\ &= \mu(\Omega) - \sup\{\mu(B) : \mathcal{A} \ni B \subset A^c\} \\ &= \mu(\Omega) - \mu(A^c) = \mu(A).\end{aligned}$$

2. Similarly, since $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 5.22 it follows that

$$\begin{aligned}\mu(A) = \mu(\Omega) - \mu(A^c) &= \mu(\Omega) - \uparrow \lim_{n\to\infty} \mu(A_n^c) \\ &=\downarrow \lim_{n\to\infty} [\mu(\Omega) - \mu(A_n^c)] =\downarrow \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

3. Suppose $A, B \in \mathcal{A}_\delta$ and $A_n, B_n \in \mathcal{A}$ such that $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \cup B_n \downarrow A \cup B$ and $A_n \cap B_n \downarrow A \cap B$ and therefore,

$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= \lim_{n\to\infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] \\ &= \lim_{n\to\infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).\end{aligned}$$

All we really need is the finite additivity of μ which can be proved as follows. Suppose that $A, B \in \mathcal{A}_\delta$ are disjoint, then $A \cap B = \emptyset$ implies $A^c \cup B^c = \Omega$. So by the strong additivity of μ on $\mathcal{A}_\sigma$ it follows that

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

from which it follows that

$$\begin{aligned}\mu(A \cup B) &= \mu(\Omega) - \mu(A^c \cap B^c) \\ &= \mu(\Omega) - [\mu(A^c) + \mu(B^c) - \mu(\Omega)] \\ &= \mu(A) + \mu(B).\end{aligned}$$

4. Since $A^c, C \in \mathcal{A}_\sigma$ we may use the strong additivity of μ on $\mathcal{A}_\sigma$ to conclude,

$$\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).$$

Because $\Omega = A^c \cup C$, and $\mu(A^c) = \mu(\Omega) - \mu(A)$, the above equation may be written as

$$\mu(\Omega) + \mu(C \setminus A) = \mu(\Omega) - \mu(A) + \mu(C)$$

which finishes the proof.

■

Notation 5.24 (Inner and outer measures) *Let $\mu : \mathcal{A} \to [0, \infty)$ be a finite premeasure extended to $\mathcal{A}_\sigma \cup \mathcal{A}_\delta$ as above. The for **any** $B \subset \Omega$ let*

$$\begin{aligned}&\mu_*(B) := \sup\{\mu(A) : \mathcal{A}_\delta \ni A \subset B\} \text{ and} \\ &\mu^*(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\}.\end{aligned}$$

*We refer to $\mu_*(B)$ and $\mu^*(B)$ as the **inner and outer** content of B respectively.*

If $B \subset \Omega$ has the same inner and outer content it is reasonable to define the measure of B as this common value. As we will see in Theorem 5.27 below, this extension becomes a σ – additive measure on a σ – algebra of subsets of Ω.

Definition 5.25 (Measurable Sets). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$. We say that $B \subset \Omega$ is **measurable** if $\mu_*(B) = \mu^*(B)$. We will denote the collection of measurable subsets of Ω by $\mathcal{B} = \mathcal{B}(\mu)$ and define $\bar{\mu} : \mathcal{B} \to [0, \mu(\Omega)]$ by*

$$\bar{\mu}(B) := \mu_*(B) = \mu^*(B) \text{ for all } B \in \mathcal{B}. \tag{5.7}$$

Remark 5.26. Observe that $\mu_*(B) = \mu^*(B)$ iff for all $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu(C) - \mu(A) < \varepsilon,$$

wherein we have used Lemma 5.23 for the first equality. Moreover we will use below that if $B \in \mathcal{B}$ and $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$, then

$$\mu(A) \le \mu_*(B) = \bar{\mu}(B) = \mu^*(B) \le \mu(C). \tag{5.8}$$

Theorem 5.27 (Finite Premeasure Extension Theorem). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^\Omega$ and $\bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(\Omega)]$ be as in Definition 5.25. Then $\mathcal{B}$ is a σ – algebra on Ω which contains $\mathcal{A}$ and $\bar{\mu}$ is a σ – additive measure on $\mathcal{B}$. Moreover, $\bar{\mu}$ is the unique measure on $\mathcal{B}$ such that $\bar{\mu}|_{\mathcal{A}} = \mu$.*

Proof. It is clear that $\mathcal{A} \subset \mathcal{B}$ and that $\mathcal{B}$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}$ for $i = 1, 2$ and $\varepsilon > 0$ is given. We may then choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon$ for $i = 1, 2$. Then with $A = A_1 \cup A_2$, $B = B_1 \cup B_2$ and $C = C_1 \cup C_2$, we have $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$. Since

$$C \setminus A = (C_1 \setminus A) \cup (C_2 \setminus A) \subset (C_1 \setminus A_1) \cup (C_2 \setminus A_2),$$

it follows from the sub-additivity of μ that with

$$\mu(C \setminus A) \le \mu(C_1 \setminus A_1) + \mu(C_2 \setminus A_2) < 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have shown that $B \in \mathcal{B}$. Hence we now know that $\mathcal{B}$ is an algebra.

Because $\mathcal{B}$ is an algebra, to verify that $\mathcal{B}$ is a σ – algebra it suffices to show that $B = \sum_{n=1}^\infty B_n \in \mathcal{B}$ whenever $\{B_n\}_{n=1}^\infty$ is a disjoint sequence in $\mathcal{B}$. To prove $B \in \mathcal{B}$, let $\varepsilon > 0$ be given and choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon 2^{-i}$ for all i. Since the $\{A_i\}_{i=1}^\infty$ are pairwise disjoint we may use Lemma 5.23 to show,

$$\begin{aligned}\sum_{i=1}^n \mu(C_i) &= \sum_{i=1}^n (\mu(A_i) + \mu(C_i \setminus A_i)) \\ &= \mu(\cup_{i=1}^n A_i) + \sum_{i=1}^n \mu(C_i \setminus A_i) \le \mu(\Omega) + \sum_{i=1}^n \varepsilon 2^{-i}.\end{aligned}$$

Passing to the limit, $n \to \infty$, in this equation then shows

$$\sum_{i=1}^\infty \mu(C_i) \le \mu(\Omega) + \varepsilon < \infty. \tag{5.9}$$

Let $B = \cup_{i=1}^\infty B_i$, $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$ and for $n \in \mathbb{N}$ let $A^n := \sum_{i=1}^n A_i \in \mathcal{A}_\delta$. Then $\mathcal{A}_\delta \ni A^n \subset B \subset C \in \mathcal{A}_\sigma$, $C \setminus A^n \in \mathcal{A}_\sigma$ and

$$C \setminus A^n = \cup_{i=1}^\infty (C_i \setminus A^n) \subset [\cup_{i=1}^n (C_i \setminus A_i)] \cup [\cup_{i=n+1}^\infty C_i] \in \mathcal{A}_\sigma.$$

Therefore, using the sub-additivity of μ on $\mathcal{A}_\sigma$ and the estimate (5.9),

$$\begin{aligned}\mu(C \setminus A^n) &\le \sum_{i=1}^n \mu(C_i \setminus A_i) + \sum_{i=n+1}^\infty \mu(C_i) \\ &\le \varepsilon + \sum_{i=n+1}^\infty \mu(C_i) \to \varepsilon \text{ as } n \to \infty.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, it follows that $B \in \mathcal{B}$ and that

$$\sum_{i=1}^n \mu(A_i) = \mu(A^n) \le \bar{\mu}(B) \le \mu(C) \le \sum_{i=1}^\infty \mu(C_i).$$

Letting $n \to \infty$ in this equation then shows,

$$\sum_{i=1}^\infty \mu(A_i) \le \bar{\mu}(B) \le \sum_{i=1}^\infty \mu(C_i). \tag{5.10}$$

On the other hand, since $A_i \subset B_i \subset C_i$, it follows (see Eq. (5.8) that

$$\sum_{i=1}^\infty \mu(A_i) \le \sum_{i=1}^\infty \bar{\mu}(B_i) \le \sum_{i=1}^\infty \mu(C_i). \tag{5.11}$$

As

$$\sum_{i=1}^\infty \mu(C_i) - \sum_{i=1}^\infty \mu(A_i) = \sum_{i=1}^\infty \mu(C_i \setminus A_i) \le \sum_{i=1}^\infty \varepsilon 2^{-i} = \varepsilon,$$

we may conclude from Eqs. (5.10) and (5.11) that

$$\left|\bar{\mu}(B) - \sum_{i=1}^\infty \bar{\mu}(B_i)\right| \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have shown $\bar{\mu}(B) = \sum_{i=1}^\infty \bar{\mu}(B_i)$. This completes the proof that $\mathcal{B}$ is a σ - algebra and that $\bar{\mu}$ is a measure on $\mathcal{B}$.

Since we really had no choice as to how to extend μ, it is to be expected that the extension is unique. You are asked to supply the details in Exercise 5.3 below. ■

Exercise 5.3. Let μ, $\bar{\mu}$, $\mathcal{A}$, and $\mathcal{B} := \mathcal{B}(\mu)$ be as in Theorem 5.27. Further suppose that $\mathcal{B}_0 \subset 2^\Omega$ is a σ – algebra such that $\mathcal{A} \subset \mathcal{B}_0 \subset \mathcal{B}$ and $\nu : \mathcal{B}_0 \to [0, \mu(\Omega)]$ is a σ – additive measure on $\mathcal{B}_0$ such that $\nu = \mu$ on $\mathcal{A}$. Show that $\nu = \bar{\mu}$ on $\mathcal{B}_0$ as well. (When $\mathcal{B}_0 = \sigma(\mathcal{A})$ this exercise is of course a consequence of Proposition 5.15. It is not necessary to use this information to complete the exercise.)

Corollary 5.28. *Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra and $\mu : \mathcal{B}_0 := \sigma(\mathcal{A}) \to [0, \mu(\Omega)]$ is a σ – additive measure. Then for every $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$;*

1. *there exists $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$ and $\varepsilon > 0$ such that $\mu(C \setminus A) < \varepsilon$ and*
2. *there exists $A \in \mathcal{A}$ such that $\mu(A \Delta B) < \varepsilon$.*

Exercise 5.4. Prove corollary 5.28 by considering $\bar{\nu}$ where $\nu := \mu|_\mathcal{A}$. **Hint:** you may find Exercise 4.3 useful here.

Theorem 5.29. *Suppose that μ is a σ – finite premeasure on an algebra $\mathcal{A}$. Then*

$$\bar{\mu}(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\} \ \forall \ B \in \sigma(\mathcal{A}) \tag{5.12}$$

defines a measure on $\sigma(\mathcal{A})$ and this measure is the unique extension of μ on $\mathcal{A}$ to a measure on $\sigma(\mathcal{A})$.

Proof. Let $\{\Omega_n\}_{n=1}^\infty \subset \mathcal{A}$ be chosen so that $\mu(\Omega_n) < \infty$ for all n and $\Omega_n \uparrow \Omega$ as $n \to \infty$ and let

$$\mu_n(A) := \mu_n(A \cap \Omega_n) \text{ for all } A \in \mathcal{A}.$$

Each μ_n is a premeasure (as is easily verified) on $\mathcal{A}$ and hence by Theorem 5.27 each μ_n has an extension, $\bar{\mu}_n$, to a measure on $\sigma(\mathcal{A})$. Since the measure $\bar{\mu}_n$ are increasing, $\bar{\mu} := \lim_{n\to\infty} \bar{\mu}_n$ is a measure which extends μ.

The proof will be completed by verifying that Eq. (5.12) holds. Let $B \in \sigma(\mathcal{A})$, $B_m = \Omega_m \cap B$ and $\varepsilon > 0$ be given. By Theorem 5.27, there exists $C_m \in \mathcal{A}_\sigma$ such that $B_m \subset C_m \subset \Omega_m$ and $\bar{\mu}(C_m \setminus B_m) = \bar{\mu}_m(C_m \setminus B_m) < \varepsilon 2^{-n}$. Then $C := \cup_{m=1}^\infty C_m \in \mathcal{A}_\sigma$ and

$$\bar{\mu}(C \setminus B) \le \bar{\mu}\left(\bigcup_{m=1}^\infty (C_m \setminus B)\right) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B) \le \sum_{m=1}^\infty \bar{\mu}(C_m \setminus B_m) < \varepsilon.$$

Thus

$$\bar{\mu}(B) \le \bar{\mu}(C) = \bar{\mu}(B) + \bar{\mu}(C \setminus B) \le \bar{\mu}(B) + \varepsilon$$

which, since $\varepsilon > 0$ is arbitrary, shows $\bar{\mu}$ satisfies Eq. (5.12). The uniqueness of the extension $\bar{\mu}$ is proved in Exercise 5.11. ■

The following slight reformulation of Theorem 5.29 can be useful.

Corollary 5.30. *Let $\mathcal{A}$ be an algebra of sets, $\{\Omega_m\}_{m=1}^\infty \subset \mathcal{A}$ is a given sequence of sets such that $\Omega_m \uparrow \Omega$ as $m \to \infty$. Let*

$$\mathcal{A}_f := \{A \in \mathcal{A} : A \subset \Omega_m \text{ for some } m \in \mathbb{N}\}.$$

Notice that $\mathcal{A}_f$ is a ring, i.e. closed under differences, intersections and unions and contains the empty set. Further suppose that $\mu : \mathcal{A}_f \to [0, \infty)$ is an additive set function such that $\mu(A_n) \downarrow 0$ for any sequence, $\{A_n\} \subset \mathcal{A}_f$ such that $A_n \downarrow \emptyset$ as $n \to \infty$. Then μ extends uniquely to a σ – finite measure on $\mathcal{A}$.

Proof. Existence. By assumption, $\mu_m := \mu|_{\mathcal{A}_{\Omega_m}} : \mathcal{A}_{\Omega_m} \to [0, \infty)$ is a premeasure on $(\Omega_m, \mathcal{A}_{\Omega_m})$ and hence by Theorem 5.29 extends to a measure μ'_m on $(\Omega_m, \sigma(\mathcal{A}_{\Omega_m}) = \mathcal{B}_{\Omega_m})$. Let $\bar{\mu}_m(B) := \mu'_m(B \cap \Omega_m)$ for all $B \in \mathcal{B}$. Then $\{\bar{\mu}_m\}_{m=1}^\infty$ is an increasing sequence of measure on $(\Omega, \mathcal{B})$ and hence $\bar{\mu} := \lim_{m\to\infty} \bar{\mu}_m$ defines a measure on $(\Omega, \mathcal{B})$ such that $\bar{\mu}|_{\mathcal{A}_f} = \mu$.

Uniqueness. If μ_1 and μ_2 are two such extensions, then $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{A}$ and therefore by Proposition 5.15 or Exercise 5.11 we know that $\mu_1(\Omega_m \cap B) = \mu_2(\Omega_m \cap B)$ for all $B \in \mathcal{B}$. We may now let $m \to \infty$ to see that in fact $\mu_1(B) = \mu_2(B)$ for all $B \in \mathcal{B}$, i.e. $\mu_1 = \mu_2$. ■

5.4 Radon Measures on $\mathbb{R}$

We say that a measure, μ, on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ is a **Radon measure** if $\mu([a, b]) < \infty$ for all $-\infty < a < b < \infty$. In this section we will give a characterization of all Radon measures on $\mathbb{R}$. We first need the following general result characterizing premeasures on an algebra generated by a semi-algebra.

Proposition 5.31. *Suppose that $\mathcal{S} \subset 2^\Omega$ is a semi-algebra, $\mathcal{A} = \mathcal{A}(\mathcal{S})$ and $\mu : \mathcal{A} \to [0, \infty]$ is a finitely additive measure. Then μ is a premeasure on $\mathcal{A}$ iff μ is countably sub-additive on $\mathcal{S}$.*

Proof. Clearly if μ is a premeasure on $\mathcal{A}$ then μ is σ - additive and hence sub-additive on $\mathcal{S}$. Because of Proposition 4.2, to prove the converse it suffices to show that the sub-additivity of μ on $\mathcal{S}$ implies the sub-additivity of μ on $\mathcal{A}$.

So suppose $A = \sum_{n=1}^\infty A_n \in \mathcal{A}$ with each $A_n \in \mathcal{A}$. By Proposition 3.25 we may write $A = \sum_{j=1}^k E_j$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_j, E_{n,i} \in \mathcal{S}$. Intersecting the identity, $A = \sum_{n=1}^\infty A_n$, with E_j implies

$$E_j = A \cap E_j = \sum_{n=1}^\infty A_n \cap E_j = \sum_{n=1}^\infty \sum_{i=1}^{N_n} E_{n,i} \cap E_j.$$

By the assumed sub-additivity of μ on $\mathcal{S}$,

$$\mu(E_j) \le \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu\left(E_{n,i} \cap E_j\right).$$

Summing this equation on j and using the finite additivity of μ shows

$$\mu(A) = \sum_{j=1}^{k} \mu(E_j) \le \sum_{j=1}^{k} \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu\left(E_{n,i} \cap E_j\right)$$
$$= \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \sum_{j=1}^{k} \mu\left(E_{n,i} \cap E_j\right) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu\left(E_{n,i}\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right).$$

■

Suppose now that μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that

$$\mu\left((a,b]\right) = F\left(b\right) - F\left(a\right) \text{ for all } -\infty < a \le b < \infty. \tag{5.13}$$

For example if $\mu\left(\mathbb{R}\right) < \infty$ we can take $F\left(x\right) = \mu\left((-\infty, x]\right)$ while if $\mu\left(\mathbb{R}\right) = \infty$ we might take

$$F\left(x\right) = \begin{cases} \mu\left((0,x]\right) & \text{if } x \ge 0 \\ -\mu\left((x,0]\right) & \text{if } x \le 0 \end{cases}.$$

The function F is uniquely determined modulo translation by a constant.

Lemma 5.32. *If μ is a Radon measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $F : \mathbb{R} \to \mathbb{R}$ is chosen so that $\mu\left((a,b]\right) = F\left(b\right) - F\left(a\right)$, then F is increasing and right continuous.*

Proof. The function F is increasing by the monotonicity of μ. To see that F is right continuous, let $b \in \mathbb{R}$ and choose $a \in (-\infty, b)$ and any sequence $\{b_n\}_{n=1}^{\infty} \subset (b, \infty)$ such that $b_n \downarrow b$ as $n \to \infty$. Since $\mu\left((a, b_1]\right) < \infty$ and $(a, b_n] \downarrow (a, b]$ as $n \to \infty$, it follows that

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

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $\lim_{y \downarrow b} F(y) = F(b)$. ■

The key result of this section is the converse to this lemma.

Theorem 5.33. *Suppose $F : \mathbb{R} \to \mathbb{R}$ is a right continuous increasing function. Then there exists a unique Radon measure, $\mu = \mu_F$, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that Eq. (5.13) holds.*

Proof. Let $\mathcal{S} := \{(a,b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$, and $\mathcal{A} = \mathcal{A}\left(\mathcal{S}\right)$ consists of those sets, $A \subset \mathbb{R}$ which may be written as finite disjoint unions of sets from $\mathcal{S}$ as in Example 3.26. Recall that $\mathcal{B}_{\mathbb{R}} = \sigma\left(\mathcal{A}\right) = \sigma\left(\mathcal{S}\right)$. Further define $F\left(\pm\infty\right) := \lim_{x \to \pm\infty} F\left(x\right)$ and let $\mu = \mu_F$ be the finitely additive measure on $(\mathbb{R}, \mathcal{A})$ described in Proposition 4.8 and Remark 4.9. To finish the proof it suffices by Theorem 5.29 to show that μ is a premeasure on $\mathcal{A} = \mathcal{A}\left(\mathcal{S}\right)$ where $\mathcal{S} := \{(a,b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}$. So in light of Proposition 5.31, to finish the proof it suffices to show μ is sub-additive on $\mathcal{S}$, i.e. we must show

$$\mu(J) \le \sum_{n=1}^{\infty} \mu(J_n). \tag{5.14}$$

where $J = \sum_{n=1}^{\infty} J_n$ with $J = (a,b] \cap \mathbb{R}$ and $J_n = (a_n, b_n] \cap \mathbb{R}$. Recall from Proposition 4.2 that the finite additivity of μ implies

$$\sum_{n=1}^{\infty} \mu(J_n) \le \mu\left(J\right). \tag{5.15}$$

We begin with the special case where $-\infty < a < b < \infty$. Our proof will be by "continuous induction." The strategy is to show $a \in \Lambda$ where

$$\Lambda := \left\{ \alpha \in [a,b] : \mu(J \cap (\alpha, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha, b]) \right\}. \tag{5.16}$$

As $b \in J$, there exists an k such that $b \in J_k$ and hence $(a_k, b_k] = (a_k, b]$ for this k. It now easily follows that $J_k \subset \Lambda$ so that Λ is not empty. To finish the proof we are going to show $\bar{a} := \inf \Lambda \in \Lambda$ and that $\bar{a} = a$.

- If $\bar{a} \notin \Lambda$, there would exist $\alpha_m \in \Lambda$ such that $\alpha_m \downarrow \bar{a}$, i.e.

$$\mu(J \cap (\alpha_m, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha_m, b]). \tag{5.17}$$

  Since $\mu(J_n \cap (\alpha_m, b]) \le \mu(J_n)$ and $\sum_{n=1}^{\infty} \mu\left(J_n\right) \le \mu\left(J\right) < \infty$ by Eq. (5.15), we may use the right continuity of F and the dominated convergence theorem for sums in order to pass to the limit as $m \to \infty$ in Eq. (5.17) to learn,

$$\mu(J \cap (\bar{a}, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\bar{a}, b]).$$

  This shows $\bar{a} \in \Lambda$ which is a contradiction to the original assumption that $\bar{a} \notin \Lambda$.
- If $\bar{a} > a$, then $\bar{a} \in J_l = (a_l, b_l]$ for some l. Letting $\alpha = a_l < \bar{a}$, we have,

$$\begin{aligned}\mu(J\cap(\alpha,b]) &= \mu(J\cap(\alpha,\bar{a}])+\mu(J\cap(\bar{a},b]) \\ &\leq \mu(J_l\cap(\alpha,\bar{a}])+\sum_{n=1}^{\infty}\mu(J_n\cap(\bar{a},b]) \\ &= \mu(J_l\cap(\alpha,\bar{a}])+\mu\left(J_l\cap(\bar{a},b]\right)+\sum_{n\neq l}\mu(J_n\cap(\bar{a},b]) \\ &= \mu(J_l\cap(\alpha,b])+\sum_{n\neq l}\mu(J_n\cap(\bar{a},b]) \\ &\leq \sum_{n=1}^{\infty}\mu(J_n\cap(\alpha,b]).\end{aligned}$$

This shows $\alpha\in\Lambda$ and $\alpha<\bar{a}$ which violates the definition of $\bar{a}$. Thus we must conclude that $\bar{a}=a$.

The hard work is now done but we still have to check the cases where $a=-\infty$ or $b=\infty$. For example, suppose that $b=\infty$ so that

$$J=(a,\infty)=\sum_{n=1}^{\infty}J_n$$

with $J_n=(a_n,b_n]\cap\mathbb{R}$. Then

$$I_M:=(a,M]=J\cap I_M=\sum_{n=1}^{\infty}J_n\cap I_M$$

and so by what we have already proved,

$$F(M)-F(a)=\mu(I_M)\leq\sum_{n=1}^{\infty}\mu(J_n\cap I_M)\leq\sum_{n=1}^{\infty}\mu(J_n).$$

Now let $M\to\infty$ in this last inequality to find that

$$\mu((a,\infty))=F(\infty)-F(a)\leq\sum_{n=1}^{\infty}\mu(J_n).$$

The other cases where $a=-\infty$ and $b\in\mathbb{R}$ and $a=-\infty$ and $b=\infty$ are handled similarly. ■

5.4.1 Lebesgue Measure

If $F(x)=x$ for all $x\in\mathbb{R}$, we denote μ_F by m and call m Lebesgue measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$.

Theorem 5.34. *Lebesgue measure m is invariant under translations, i.e. for $B\in\mathcal{B}_{\mathbb{R}}$ and $x\in\mathbb{R}$,*

$$m(x+B)=m(B). \tag{5.18}$$

Lebesgue measure, m, is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0,1])=1$ and Eq. (5.18) holds for $B\in\mathcal{B}_{\mathbb{R}}$ and $x\in\mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B)=|\lambda|\,m(B) \tag{5.19}$$

where $\lambda\in\mathbb{R}$, $B\in\mathcal{B}_{\mathbb{R}}$ and $\lambda B:=\{\lambda x: x\in B\}$.

Proof. Let $m_x(B):=m(x+B)$, then one easily shows that m_x is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a,b])=b-a$ for all $a<b$. Therefore, $m_x=m$ by the uniqueness assertion in Exercise 5.11. For the converse, suppose that m is translation invariant and $m((0,1])=1$. Given $n\in\mathbb{N}$, we have

$$(0,1]=\cup_{k=1}^{n}(\frac{k-1}{n},\frac{k}{n}]=\cup_{k=1}^{n}\left(\frac{k-1}{n}+(0,\frac{1}{n}]\right).$$

Therefore,

$$\begin{aligned}1=m((0,1]) &= \sum_{k=1}^{n}m\left(\frac{k-1}{n}+(0,\frac{1}{n}]\right) \\ &= \sum_{k=1}^{n}m((0,\frac{1}{n}])=n\cdot m((0,\frac{1}{n}]).\end{aligned}$$

That is to say

$$m((0,\frac{1}{n}])=1/n.$$

Similarly, $m((0,\frac{l}{n}])=l/n$ for all $l,n\in\mathbb{N}$ and therefore by the translation invariance of m,

$$m((a,b])=b-a\text{ for all }a,b\in\mathbb{Q}\text{ with }a<b.$$

Finally for $a,b\in\mathbb{R}$ such that $a<b$, choose $a_n,b_n\in\mathbb{Q}$ such that $b_n\downarrow b$ and $a_n\uparrow a$, then $(a_n,b_n]\downarrow(a,b]$ and thus

$$m((a,b])=\lim_{n\to\infty}m((a_n,b_n])=\lim_{n\to\infty}(b_n-a_n)=b-a,$$

i.e. m is Lebesgue measure. To prove Eq. (5.19) we may assume that $\lambda\neq 0$ since this case is trivial to prove. Now let $m_\lambda(B):=|\lambda|^{-1}m(\lambda B)$. It is easily checked that m_λ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

$$m_\lambda((a,b])=\lambda^{-1}m\left((\lambda a,\lambda b]\right)=\lambda^{-1}(\lambda b-\lambda a)=b-a$$

if $\lambda>0$ and

$$m_\lambda((a,b])=|\lambda|^{-1}m\left([\lambda b,\lambda a)\right)=-|\lambda|^{-1}(\lambda b-\lambda a)=b-a$$

if $\lambda<0$. Hence $m_\lambda=m$. ■

5.5 A Discrete Kolmogorov's Extension Theorem

For this section, let S be a finite or countable set (we refer to S as **state space**), $\Omega := S^\infty := S^{\mathbb{N}}$ (think of $\mathbb{N}$ as time and Ω as **path space**)

$$\mathcal{A}_n := \{B \times \Omega : B \subset S^n\} \text{ for all } n \in \mathbb{N},$$

$\mathcal{A} := \cup_{n=1}^\infty \mathcal{A}_n$, and $\mathcal{B} := \sigma(\mathcal{A})$. We call the elements, $A \subset \Omega$, the **cylinder subsets of** Ω. Notice that $A \subset \Omega$ is a cylinder set iff there exists $n \in \mathbb{N}$ and $B \subset S^n$ such that

$$A = B \times \Omega := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Also observe that we may write A as $A = B' \times \Omega$ where $B' = B \times S^k \subset S^{n+k}$ for any $k \geq 0$.

Exercise 5.5. Show;

1. $\mathcal{A}_n$ is a σ – algebra for each $n \in \mathbb{N}$,
2. $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ for all n, and
3. $\mathcal{A} \subset 2^\Omega$ is an algebra of subsets of Ω. (In fact, you might show that $\mathcal{A} = \cup_{n=1}^\infty \mathcal{A}_n$ is an algebra whenever $\{\mathcal{A}_n\}_{n=1}^\infty$ is an increasing sequence of algebras.)

Lemma 5.35 (Baby Tychonov Theorem). *Suppose $\{C_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of **non-empty** cylinder sets. Further assume there exists $N_n \in \mathbb{N}$ and $B_n \subset\subset S^{N_n}$ such that $C_n = B_n \times \Omega$. (This last assumption is vacuous when S is a finite set. Recall that we write $\Lambda \subset\subset A$ to indicate that Λ is a finite subset of A.) Then $\cap_{n=1}^\infty C_n \neq \emptyset$.*

Proof. Since $C_{n+1} \subset C_n$, if $N_n > N_{n+1}$, we would have $B_{n+1} \times S^{N_{n+1}-N_n} \subset B_n$. If S is an infinite set this would imply B_n is an infinite set and hence we must have $N_{n+1} \geq N_n$ for all n when $\#(S) = \infty$. On the other hand, if S is a finite set, we can always replace B_{n+1} by $B_{n+1} \times S^k$ for some appropriate k and arrange it so that $N_{n+1} \geq N_n$ for all n. So from now we assume that $N_{n+1} \geq N_n$.

Case 1. $\lim_{n\to\infty} N_n < \infty$ in which case there exists some $N \in \mathbb{N}$ such that $N_n = N$ for all large n. Thus for large N, $C_n = B_n \times \Omega$ with $B_n \subset\subset S^N$ and $B_{n+1} \subset B_n$ and hence $\#(B_n) \downarrow$ as $n \to \infty$. By assumption, $\lim_{n\to\infty} \#(B_n) \neq 0$ and therefore $\#(B_n) = k > 0$ for all n large. It then follows that there exists $n_0 \in \mathbb{N}$ such that $B_n = B_{n_0}$ for all $n \geq n_0$. Therefore $\cap_{n=1}^\infty C_n = B_{n_0} \times \Omega \neq \emptyset$.

Case 2. $\lim_{n\to\infty} N_n = \infty$. By assumption, there exists $\omega(n) = (\omega_1(n), \omega_2(n), \dots) \in \Omega$ such that $\omega(n) \in C_n$ for all n. Moreover, since $\omega(n) \in C_n \subset C_k$ for all $k \leq n$, it follows that

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } n \geq k \tag{5.20}$$

and as B_k is a finite set $\{\omega_i(n)\}_{n=1}^\infty$ must be a finite set for all $1 \leq i \leq N_k$. As $N_k \to \infty$ as $k \to \infty$ it follows that $\{\omega_i(n)\}_{n=1}^\infty$ is a finite set for all $i \in \mathbb{N}$. Using this observation, we may find, $s_1 \in S$ and an infinite subset, $\Gamma_1 \subset \mathbb{N}$ such that $\omega_1(n) = s_1$ for all $n \in \Gamma_1$. Similarly, there exists $s_2 \in S$ and an infinite set, $\Gamma_2 \subset \Gamma_1$, such that $\omega_2(n) = s_2$ for all $n \in \Gamma_2$. Continuing this procedure inductively, there exists (for all $j \in \mathbb{N}$) infinite subsets, $\Gamma_j \subset \mathbb{N}$ and points $s_j \in S$ such that $\Gamma_1 \supset \Gamma_2 \supset \Gamma_3 \supset \dots$ and $\omega_j(n) = s_j$ for all $n \in \Gamma_j$.

We are now going to complete the proof by showing $s := (s_1, s_2, \dots) \in \cap_{n=1}^\infty C_n$. By the construction above, for all $N \in \mathbb{N}$ we have

$$(\omega_1(n), \dots, \omega_N(n)) = (s_1, \dots, s_N) \text{ for all } n \in \Gamma_N.$$

Taking $N = N_k$ and $n \in \Gamma_{N_k}$ with $n \geq k$, we learn from Eq. (5.20) that

$$(s_1, \dots, s_{N_k}) = (\omega_1(n), \dots, \omega_{N_k}(n)) \in B_k.$$

But this is equivalent to showing $s \in C_k$. Since $k \in \mathbb{N}$ was arbitrary it follows that $s \in \cap_{n=1}^\infty C_n$. ■

Let $\bar{S} := S$ is S is a finite set and $\bar{S} = S \cup \{\infty\}$ if S is an infinite set. Here, ∞, is simply another point not in S which we call infinity Let $\{x_n\}_{n=1}^\infty \subset \bar{S}$ be a sequence, then we way $\lim_{n\to\infty} x_n = \infty$ if for every $A \subset\subset S$, $x_n \notin A$ for almost all n and we say that $\lim_{n\to\infty} x_n = s \in S$ if $x_n = s$ for almost all n. For example this is the usual notion of convergence for $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ and $\bar{S} = S \cup \{0\} \subset [0,1]$, where 0 is playing the role of infinity here. Observe that either $\lim_{n\to\infty} x_n = \infty$ or there exists a finite subset $F \subset S$ such that $x_n \in F$ infinitely often. Moreover, there must be some point, $s \in F$ such that $x_n = s$ infinitely often. Thus if we let $\{n_1 < n_2 < \dots\} \subset \mathbb{N}$ be chosen such that $x_{n_k} = s$ for all k, then $\lim_{k\to\infty} x_{n_k} = s$. Thus we have shown that every sequence in $\bar{S}$ has a convergent subsequence.

Lemma 5.36 (Baby Tychonov Theorem I.). *Let $\bar{\Omega} := \bar{S}^{\mathbb{N}}$ and $\{\omega(n)\}_{n=1}^\infty$ be a sequence in $\bar{\Omega}$. Then there is a subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$ such that $\lim_{k\to\infty} \omega(n_k)$ exists in $\bar{\Omega}$ by which we mean, $\lim_{k\to\infty} \omega_i(n_k)$ exists in $\bar{S}$ for all $i \in \mathbb{N}$.*

Proof. This follows by the usual cantor's diagonalization argument. Indeed, let $\{n_k^1\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ be chosen so that $\lim_{k\to\infty} \omega_1(n_k^1) = s_1 \in \bar{S}$ exists. Then choose $\{n_k^2\}_{k=1}^\infty \subset \{n_k^1\}_{k=1}^\infty$ so that $\lim_{k\to\infty} \omega_2(n_k^2) = s_2 \in \bar{S}$ exists. Continue on this way to inductively choose

$$\{n_k^1\}_{k=1}^\infty \supset \{n_k^2\}_{k=1}^\infty \supset \dots \supset \{n_k^l\}_{k=1}^\infty \supset \dots$$

such that $\lim_{k\to\infty} \omega_l(n_k^l) = s_l \in \bar{S}$. The subsequence, $\{n_k\}_{k=1}^\infty$ of $\{n\}_{n=1}^\infty$, may now be defined by, $n_k = n_k^k$. ■

Corollary 5.37 (Baby Tychonov Theorem II.). *Suppose that $\{F_n\}_{n=1}^\infty \subset \bar{\Omega}$ is decreasing sequence of non-empty sets which are closed under taking sequential limits, then $\cap_{n=1}^\infty F_n \neq \emptyset$.*

Proof. Since $F_n \neq \emptyset$ there exists $\omega(n) \in F_n$ for all n. Using Lemma 5.36, there exists $\{n_k\}_{k=1}^\infty \subset \{n\}_{n=1}^\infty$ such that $\omega := \lim_{k\to\infty} \omega(n_k)$ exits in $\bar{\Omega}$. Since $\omega(n_k) \in F_n$ for all $k \geq n$, it follows that $\omega \in F_n$ for all n, i.e. $\omega \in \cap_{n=1}^\infty F_n$ and hence $\cap_{n=1}^\infty F_n \neq \emptyset$. ■

Example 5.38. Suppose that $1 \leq N_1 < N_2 < N_3 < \dots$, $F_n = K_n \times \Omega$ with $K_n \subset\subset S^{N_n}$ such that $\{F_n\}_{n=1}^\infty \subset \Omega$ is a decreasing sequence of non-empty sets. Then $\cap_{n=1}^\infty F_n \neq \emptyset$. To prove this, let $\bar{F}_n := K_n \times \bar{\Omega}$ in which case $\bar{F}_n$ are non – empty sets closed under taking limits. Therefore by Corollary 5.37, $\cap_n \bar{F}_n \neq \emptyset$. This completes the proof since it is easy to check that $\cap_{n=1}^\infty F_n = \cap_n \bar{F}_n \neq \emptyset$.

Corollary 5.39. *If S is a finite set and $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

Proof. This follows directly from Example 5.38 since necessarily, $A_n = K_n \times \Omega$, for some $K_n \subset\subset S^{N_n}$. ■

Theorem 5.40 (Kolmogorov's Extension Theorem I.). *Let us continue the notation above with the further assumption that S is a finite set. Then* ***every*** *finitely additive probability measure, $P : \mathcal{A} \to [0,1]$, has a unique extension to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 5.27, it suffices to show $\lim_{n\to\infty} P(A_n) = 0$ whenever $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ with $A_n \downarrow \emptyset$. However, by Lemma 5.35 with $C_n = A_n$, $A_n \in \mathcal{A}$ and $A_n \downarrow \emptyset$, we must have that $A_n = \emptyset$ for a.a. n and in particular $P(A_n) = 0$ for a.a. n. This certainly implies $\lim_{n\to\infty} P(A_n) = 0$. ■

For the next three exercises, suppose that S is a finite set and continue the notation from above. Further suppose that $P : \sigma(\mathcal{A}) \to [0,1]$ is a probability measure and for $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, let

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \quad (5.21)$$

Exercise 5.6 (Consistency Conditions). If p_n is defined as above, show:

1. $\sum_{s\in S} p_1(s) = 1$ and
2. for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$,

$$p_n(s_1, \dots, s_n) = \sum_{s\in S} p_{n+1}(s_1, \dots, s_n, s).$$

Exercise 5.7 (Converse to 5.6). Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : S^n \to [0,1]$ such that the consistency conditions in Exercise 5.6 hold. Then there exists a unique probability measure, P on $\sigma(\mathcal{A})$ such that Eq. (5.21) holds for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$.

Example 5.41 (Existence of iid simple R.V.s). Suppose now that $q : S \to [0,1]$ is a function such that $\sum_{s\in S} q(s) = 1$. Then there exists a unique probability measure P on $\sigma(\mathcal{A})$ such that, for all $n \in \mathbb{N}$ and $(s_1, \dots, s_n) \in S^n$, we have

$$P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}) = q(s_1) \dots q(s_n).$$

This is a special case of Exercise 5.7 with $p_n(s_1, \dots, s_n) := q(s_1) \dots q(s_n)$.

Theorem 5.42 (Kolmogorov's Extension Theorem II). *Suppose now that S is countably infinite set and $P : \mathcal{A} \to [0,1]$ is a finitely additive measure such that $P|_{\mathcal{A}_n}$ is a σ – additive measure for each $n \in \mathbb{N}$. Then P extends uniquely to a probability measure on $\mathcal{B} := \sigma(\mathcal{A})$.*

Proof. From Theorem 5.27 it suffice to show; if $\{A_m\}_{m=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$, then $\cap_{m=1}^\infty A_m \neq \emptyset$. You are asked to verify this property of P in the next couple of exercises. ■

For the next couple of exercises the hypothesis of Theorem 5.42 are to be assumed.

Exercise 5.8. Show for each $n \in \mathbb{N}$, $A \in \mathcal{A}_n$, and $\varepsilon > 0$ are given. Show there exists $F \in \mathcal{A}_n$ such that $F \subset A$, $F = K \times \Omega$ with $K \subset\subset S^n$, and $P(A \setminus F) < \varepsilon$.

Exercise 5.9. Let $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ be a decreasing sequence of subsets such that $\varepsilon := \inf_m P(A_m) > 0$. Using Exercise 5.8, choose $F_m = K_m \times \Omega \subset A_m$ with $K_m \subset\subset S^{N_n}$ and $P(A_m \setminus F_m) \leq \varepsilon/2^{m+1}$. Further define $C_m := F_1 \cap \dots \cap F_m$ for each m. Show;

1. Show $A_m \setminus C_m \subset (A_1 \setminus F_1) \cup (A_2 \setminus F_2) \cup \dots \cup (A_m \setminus F_m)$ and use this to conclude that $P(A_m \setminus C_m) \leq \varepsilon/2$.
2. Conclude C_m is not empty for m.
3. Use Lemma 5.35 to conclude that $\emptyset \neq \cap_{m=1}^\infty C_m \subset \cap_{m=1}^\infty A_m$.

Exercise 5.10. Convince yourself that the results of Exercise 5.6 and 5.7 are valid when S is a countable set. (See Example 4.6.)

In summary, the main result of this section states, to any sequence of functions, $p_n : S^n \to [0,1]$, such that $\sum_{\lambda\in S^n} p_n(\lambda) = 1$ and $\sum_{s\in S} p_{n+1}(\lambda, s) = p_n(\lambda)$ for all n and $\lambda \in S^n$, there exists a unique probability measure, P, on $\mathcal{B} := \sigma(\mathcal{A})$ such that

$$P(B \times \Omega) = \sum_{\lambda\in B} p_n(\lambda) \ \forall\ B \subset S^n \text{ and } n \in \mathbb{N}.$$

Example 5.43 (Markov Chain Probabilities). Let S be a finite or at most countable state space and $p : S \times S \to [0,1]$ be a **Markov kernel,** i.e.

$$\sum_{y\in S} p(x,y) = 1 \text{ for all } x\in S. \tag{5.22}$$

Also let $\pi : S \to [0,1]$ be a probability function, i.e. $\sum_{x\in S}\pi(x) = 1$. We now take

$$\Omega := S^{\mathbb{N}_0} = \{\omega = (s_0, s_1, \dots) : s_j \in S\}$$

and let $X_n : \Omega \to S$ be given by

$$X_n(s_0, s_1, \dots) = s_n \text{ for all } n \in \mathbb{N}_0.$$

Then there exists a unique probability measure, P_π, on $\sigma(\mathcal{A})$ such that

$$P_\pi(X_0 = x_0, \dots, X_n = x_n) = \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

for all $n \in \mathbb{N}_0$ and $x_0, x_1, \dots, x_n \in S$. To see such a measure exists, we need only verify that

$$p_n(x_0, \dots, x_n) := \pi(x_0)\, p(x_0, x_1) \dots p(x_{n-1}, x_n)$$

verifies the hypothesis of Exercise 5.6 taking into account a shift of the n – index.

5.6 Appendix: Regularity and Uniqueness Results*

The goal of this appendix it to approximating measurable sets from inside and outside by classes of sets which are relatively easy to understand. Our first few results are already contained in Carathoédory's existence of measures proof. Nevertheless, we state these results again and give another somewhat independent proof.

Theorem 5.44 (Finite Regularity Result). *Suppose $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} = \sigma(\mathcal{A})$ and $\mu : \mathcal{B} \to [0,\infty)$ is a finite measure, i.e. $\mu(\Omega) < \infty$. Then for every $\varepsilon > 0$ and $B \in \mathcal{B}$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$.*

Proof. Let $\mathcal{B}_0$ denote the collection of $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there here exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$. It is now clear that $\mathcal{A} \subset \mathcal{B}_0$ and that $\mathcal{B}_0$ is closed under complementation. Now suppose that $B_i \in \mathcal{B}_0$ for $i = 1, 2, \dots$ and $\varepsilon > 0$ is given. By assumption there exists $A_i \in \mathcal{A}_\delta$ and $C_i \in \mathcal{A}_\sigma$ such that $A_i \subset B_i \subset C_i$ and $\mu(C_i \setminus A_i) < 2^{-i}\varepsilon$.

Let $A := \cup_{i=1}^\infty A_i$, $A^N := \cup_{i=1}^N A_i \in \mathcal{A}_\delta$, $B := \cup_{i=1}^\infty B_i$, and $C := \cup_{i=1}^\infty C_i \in \mathcal{A}_\sigma$. Then $A^N \subset A \subset B \subset C$ and

$$C \setminus A = [\cup_{i=1}^\infty C_i] \setminus A = \cup_{i=1}^\infty [C_i \setminus A] \subset \cup_{i=1}^\infty [C_i \setminus A_i].$$

Therefore,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^\infty [C_i \setminus A]) \le \sum_{i=1}^\infty \mu(C_i \setminus A) \le \sum_{i=1}^\infty \mu(C_i \setminus A_i) < \varepsilon.$$

Since $C \setminus A^N \downarrow C \setminus A$, it also follows that $\mu\left(C \setminus A^N\right) < \varepsilon$ for sufficiently large N and this shows $B = \cup_{i=1}^\infty B_i \in \mathcal{B}_0$. Hence $\mathcal{B}_0$ is a sub-σ-algebra of $\mathcal{B} = \sigma(\mathcal{A})$ which contains $\mathcal{A}$ which shows $\mathcal{B}_0 = \mathcal{B}$. ■

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a σ – finite measure defined next.

Definition 5.45. *Suppose Ω is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^\Omega$ and $\mu : \mathcal{B} \to [0,\infty]$ is a function. The function μ is σ – finite on $\mathcal{E}$ if there exists $E_n \in \mathcal{E}$ such that $\mu(E_n) < \infty$ and $\Omega = \cup_{n=1}^\infty E_n$. If $\mathcal{B}$ is a σ – algebra and μ is a measure on $\mathcal{B}$ which is σ – finite on $\mathcal{B}$ we will say $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space.*

The reader should check that if μ is a finitely additive measure on an algebra, $\mathcal{B}$, then μ is σ – finite on $\mathcal{B}$ iff there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$.

Corollary 5.46 (σ – Finite Regularity Result). *Theorem 5.44 continues to hold under the weaker assumption that $\mu : \mathcal{B} \to [0,\infty]$ is a measure which is σ – finite on $\mathcal{A}$.*

Proof. Let $\Omega_n \in \mathcal{A}$ such that $\cup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all n.Since $A \in \mathcal{B} \to \mu_n(A) := \mu(\Omega_n \cap A)$ is a finite measure on $A \in \mathcal{B}$ for each n, by Theorem 5.44, for every $B \in \mathcal{B}$ there exists $C_n \in \mathcal{A}_\sigma$ such that $B \subset C_n$ and $\mu(\Omega_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n}\varepsilon$. Now let $C := \cup_{n=1}^\infty [\Omega_n \cap C_n] \in \mathcal{A}_\sigma$ and observe that $B \subset C$ and

$$\begin{aligned}\mu(C \setminus B) &= \mu(\cup_{n=1}^\infty ([\Omega_n \cap C_n] \setminus B)) \\ &\le \sum_{n=1}^\infty \mu([\Omega_n \cap C_n] \setminus B) = \sum_{n=1}^\infty \mu(\Omega_n \cap [C_n \setminus B]) < \varepsilon.\end{aligned}$$

Applying this result to B^c shows there exists $D \in \mathcal{A}_\sigma$ such that $B^c \subset D$ and

$$\mu(B \setminus D^c) = \mu(D \setminus B^c) < \varepsilon.$$

So if we let $A := D^c \in \mathcal{A}_\delta$, then $A \subset B \subset C$ and

$$\mu(C \setminus A) = \mu([B \setminus A] \cup [(C \setminus B) \setminus A]) \le \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon$$

and the result is proved. ■

Exercise 5.11. Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and μ and ν are two measures on $\mathcal{B} = \sigma(\mathcal{A})$.

a. Suppose that μ and ν are finite measures such that $\mu = \nu$ on $\mathcal{A}$. Show $\mu = \nu$.
b. Generalize the previous assertion to the case where you only assume that μ and ν are σ – finite on $\mathcal{A}$.

Corollary 5.47. *Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and $\mu : \mathcal{B} = \sigma(\mathcal{A}) \to [0,\infty]$ is a measure which is σ – finite on $\mathcal{A}$. Then for all $B \in \mathcal{B}$, there exists $A \in \mathcal{A}_{\delta\sigma}$ and $C \in \mathcal{A}_{\sigma\delta}$ such that $A \subset B \subset C$ and $\mu(C \setminus A) = 0$.*

Proof. By Theorem 5.44, given $B \in \mathcal{B}$, we may choose $A_n \in \mathcal{A}_\delta$ and $C_n \in \mathcal{A}_\sigma$ such that $A_n \subset B \subset C_n$ and $\mu(C_n \setminus B) \le 1/n$ and $\mu(B \setminus A_n) \le 1/n$. By replacing A_N by $\cup_{n=1}^N A_n$ and C_N by $\cap_{n=1}^N C_n$, we may assume that $A_n \uparrow$ and $C_n \downarrow$ as n increases. Let $A = \cup A_n \in \mathcal{A}_{\delta\sigma}$ and $C = \cap C_n \in \mathcal{A}_{\sigma\delta}$, then $A \subset B \subset C$ and

$$\begin{aligned}\mu(C \setminus A) &= \mu(C \setminus B) + \mu(B \setminus A) \le \mu(C_n \setminus B) + \mu(B \setminus A_n) \\ &\le 2/n \to 0 \text{ as } n \to \infty.\end{aligned}$$

■

Exercise 5.12. Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma(\{\text{open subsets of } \mathbb{R}^n\})$ be the Borel σ – algebra on $\mathbb{R}^n$ and μ be a probability measure on $\mathcal{B}$. Further, let $\mathcal{B}_0$ denote those sets $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there exists $F \subset B \subset V$ such that F is closed, V is open, and $\mu(V \setminus F) < \varepsilon$. Show:

1. $\mathcal{B}_0$ contains all closed subsets of $\mathcal{B}$. **Hint:** given a closed subset, $F \subset \mathbb{R}^n$ and $k \in \mathbb{N}$, let $V_k := \cup_{x \in F} B(x, 1/k)$, where $B(x,\delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}$. Show, $V_k \downarrow F$ as $k \to \infty$.
2. Show $\mathcal{B}_0$ is a σ – algebra and use this along with the first part of this exercise to conclude $\mathcal{B} = \mathcal{B}_0$. **Hint:** follow closely the method used in the first step of the proof of Theorem 5.44.
3. Show for every $\varepsilon > 0$ and $B \in \mathcal{B}$, there exist a compact subset, $K \subset \mathbb{R}^n$, such that $K \subset B$ and $\mu(B \setminus K) < \varepsilon$. **Hint:** take $K := F \cap \{x \in \mathbb{R}^n : |x| \le n\}$ for some sufficiently large n.

5.7 Appendix: Completions of Measure Spaces*

Definition 5.48. *A set $E \subset \Omega$ is a **null** set if $E \in \mathcal{B}$ and $\mu(E) = 0$. If P is some "property" which is either true or false for each $x \in \Omega$, we will use the terminology P a.e. (to be read P almost everywhere) to mean*

$$E := \{x \in \Omega : P \text{ is false for } x\}$$

is a null set. For example if f and g are two measurable functions on $(\Omega, \mathcal{B}, \mu)$, $f = g$ a.e. means that $\mu(f \neq g) = 0$.

Definition 5.49. *A measure space $(\Omega, \mathcal{B}, \mu)$ is **complete** if every subset of a null set is in $\mathcal{B}$, i.e. for all $F \subset \Omega$ such that $F \subset E \in \mathcal{B}$ with $\mu(E) = 0$ implies that $F \in \mathcal{B}$.*

Proposition 5.50 (Completion of a Measure). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space. Set*

$$\begin{aligned}\mathcal{N} = \mathcal{N}^\mu &:= \{N \subset \Omega : \exists\, F \in \mathcal{B} \text{ such that } N \subset F \text{ and } \mu(F) = 0\}, \\ \mathcal{B} = \bar{\mathcal{B}}^\mu &:= \{A \cup N : A \in \mathcal{B} \text{ and } N \in \mathcal{N}\} \text{ and} \\ \bar{\mu}(A \cup N) &:= \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N},\end{aligned}$$

*see Fig. 5.2. Then $\bar{\mathcal{B}}$ is a σ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{B}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{B}}$ which extends μ on $\mathcal{B}$, and $(\Omega, \bar{\mathcal{B}}, \bar{\mu})$ is complete measure space. The σ-algebra, $\bar{\mathcal{B}}$, is called the **completion** of $\mathcal{B}$ relative to μ and $\bar{\mu}$, is called the **completion of** μ.*

Proof. Clearly $\Omega, \emptyset \in \bar{\mathcal{B}}$. Let $A \in \mathcal{B}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{B}$ such

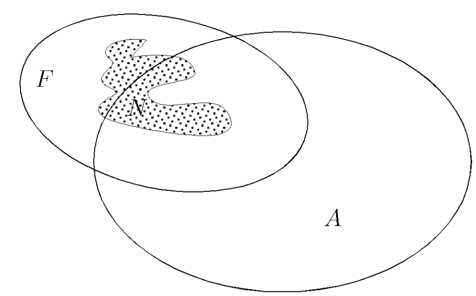


Fig. 5.2. Completing a σ – algebra.

that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$\begin{aligned}(A \cup N)^c &= A^c \cap N^c = A^c \cap (F \setminus N \cup F^c) \\ &= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]\end{aligned}$$

where $[A^c \cap (F \setminus N)] \in \mathcal{N}$ and $[A^c \cap F^c] \in \mathcal{B}$. Thus $\bar{\mathcal{B}}$ is closed under complements. If $A_i \in \mathcal{B}$ and $N_i \subset F_i \in \mathcal{B}$ such that $\mu(F_i) = 0$ then $\cup(A_i \cup N_i) = (\cup A_i) \cup (\cup N_i) \in \bar{\mathcal{B}}$ since $\cup A_i \in \mathcal{B}$ and $\cup N_i \subset \cup F_i$ and

$\mu(\cup F_i) \leq \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{B}}$ is a σ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{B}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \leq \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \leq \mu(A)$ so that $\mu(A) = \mu(B)$ and hence $\bar{\mu}(A \cup N) := \mu(A)$ is well defined. It is left as an exercise to show $\bar{\mu}$ is a measure, i.e. that it is countable additive. ■

5.8 Appendix Monotone Class Theorems*

This appendix may be safely skipped!

Definition 5.51 (Montone Class). *$\mathcal{C} \subset 2^{\Omega}$ is a* ***monotone class*** *if it is closed under countable increasing unions and countable decreasing intersections.*

Lemma 5.52 (Monotone Class Theorem*). *Suppose $\mathcal{A} \subset 2^{\Omega}$ is an algebra and $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$. Then $\mathcal{C} = \sigma(\mathcal{A})$.*

Proof. For $C \in \mathcal{C}$ let

$$\mathcal{C}(C) = \{B \in \mathcal{C} : C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}\},$$

then $\mathcal{C}(C)$ is a monotone class. Indeed, if $B_n \in \mathcal{C}(C)$ and $B_n \uparrow B$, then $B_n^c \downarrow B^c$ and so

$$\begin{aligned} &\mathcal{C} \ni C \cap B_n \uparrow C \cap B \\ &\mathcal{C} \ni C \cap B_n^c \downarrow C \cap B^c \text{ and} \\ &\mathcal{C} \ni B_n \cap C^c \uparrow B \cap C^c. \end{aligned}$$

Since $\mathcal{C}$ is a monotone class, it follows that $C \cap B, C \cap B^c, B \cap C^c \in \mathcal{C}$, i.e. $B \in \mathcal{C}(C)$. This shows that $\mathcal{C}(C)$ is closed under increasing limits and a similar argument shows that $\mathcal{C}(C)$ is closed under decreasing limits. Thus we have shown that $\mathcal{C}(C)$ is a monotone class for all $C \in \mathcal{C}$. If $A \in \mathcal{A} \subset \mathcal{C}$, then $A \cap B, A \cap B^c, B \cap A^c \in \mathcal{A} \subset \mathcal{C}$ for all $B \in \mathcal{A}$ and hence it follows that $\mathcal{A} \subset \mathcal{C}(A) \subset \mathcal{C}$. Since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$, we conclude that $\mathcal{C}(A) = \mathcal{C}$ for any $A \in \mathcal{A}$. Let $B \in \mathcal{C}$ and notice that $A \in \mathcal{C}(B)$ happens iff $B \in \mathcal{C}(A)$. This observation and the fact that $\mathcal{C}(A) = \mathcal{C}$ for all $A \in \mathcal{A}$ implies $\mathcal{A} \subset \mathcal{C}(B) \subset \mathcal{C}$ for all $B \in \mathcal{C}$. Again since $\mathcal{C}$ is the smallest monotone class containing $\mathcal{A}$ and $\mathcal{C}(B)$ is a monotone class we conclude that $\mathcal{C}(B) = \mathcal{C}$ for all $B \in \mathcal{C}$. That is to say, if $A, B \in \mathcal{C}$ then $A \in \mathcal{C} = \mathcal{C}(B)$ and hence $A \cap B$, $A \cap B^c$, $A^c \cap B \in \mathcal{C}$. So $\mathcal{C}$ is closed under complements (since $\Omega \in \mathcal{A} \subset \mathcal{C}$) and finite intersections and increasing unions from which it easily follows that $\mathcal{C}$ is a σ – algebra. ■

6

Random Variables

Notation 6.1 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E)|E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y|f^{-1}(A) \in \mathcal{G}\}.$$

Definition 6.2. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

The following results will be used frequently (often without further reference) in the sequel.

Lemma 6.3 (A key measurability lemma). *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$, then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})). \tag{6.1}$$

In particular, if $A \subset Y$ then

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{6.2}$$

(Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.)

Proof. Since $\mathcal{E} \subset \sigma(\mathcal{E})$, it follows that $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. Moreover, by Exercise 6.1 below, $f^{-1}(\sigma(\mathcal{E}))$ is a σ – algebra and therefore,

$$\sigma(f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma(\mathcal{E})).$$

To finish the proof we must show $f^{-1}(\sigma(\mathcal{E})) \subset \sigma(f^{-1}(\mathcal{E}))$, i.e. that $f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))$ for all $B \in \sigma(\mathcal{E})$. To do this we follow the usual measure theoretic mantra, namely let

$$\mathcal{M} := \left\{B \subset Y : f^{-1}(B) \in \sigma(f^{-1}(\mathcal{E}))\right\} = f_*\sigma(f^{-1}(\mathcal{E})).$$

We will now finish the proof by showing $\sigma(\mathcal{E}) \subset \mathcal{M}$. This is easily achieved by observing that $\mathcal{M}$ is a σ – algebra (see Exercise 6.1) which contains $\mathcal{E}$ and therefore $\sigma(\mathcal{E}) \subset \mathcal{M}$.

Equation (6.2) is a special case of Eq. (6.1). Indeed, $f = i_A : A \to X$ we have

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

■

Exercise 6.1. If $f : X \to Y$ is a function and $\mathcal{F} \subset 2^Y$ and $\mathcal{B} \subset 2^X$ are σ – algebras (algebras), then $f^{-1}\mathcal{F}$ and $f_*\mathcal{B}$ are σ – algebras (algebras).

Example 6.4. Let $\mathcal{E} = \{(a, b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel σ – field on $\mathbb{R}$. Then

$$\mathcal{E}_{(0,1]} = \{(a, b] : 0 \le a < b \le 1\}$$

and we have

$$\mathcal{B}_{(0,1]} = \sigma\left(\mathcal{E}_{(0,1]}\right).$$

In particular, if $A \in \mathcal{B}$ such that $A \subset (0, 1]$, then $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

6.1 Measurable Functions

Definition 6.5. *A* ***measurable space*** *is a pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ – algebra on X.*

To motivate the notion of a measurable function, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$ is a function. Roughly speaking, we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i \mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a, b]) \in \mathcal{M}$ for all $a < b$. Because of Corollary 6.11 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

Definition 6.6. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *of more precisely, $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable, if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$, i.e. if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{F}$.*

Remark 6.7. Let $f : X \to Y$ be a function. Given a σ – algebra $\mathcal{F} \subset 2^Y$, the σ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest σ – algebra on X such that f is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a σ - algebra on X then

$$\mathcal{F} = f_*\mathcal{M} = \{A \in 2^Y|f^{-1}(A) \in \mathcal{M}\}$$

is the largest σ – algebra on Y such that f is $(\mathcal{M}, \mathcal{F})$ - measurable.

Example 6.8 (Indicator Functions). Let $(X,\mathcal{M})$ be a measurable space and $A \subset X$. Then 1_A is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{M}$. Indeed, $1_A^{-1}(W)$ is either $\emptyset$, X, A or A^c for any $W \subset \mathbb{R}$ with $1_A^{-1}(\{1\}) = A$.

Example 6.9. Suppose $f : X \to Y$ with Y being a finite or countable set and $\mathcal{F} = 2^Y$. Then f is measurable iff $f^{-1}(\{y\}) \in \mathcal{M}$ for all $y \in Y$.

Proposition 6.10. *Suppose that $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ are measurable spaces and further assume $\mathcal{E} \subset \mathcal{F}$ generates $\mathcal{F}$, i.e. $\mathcal{F} = \sigma(\mathcal{E})$. Then a map, $f : X \to Y$ is measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.*

Proof. If f is $\mathcal{M}/\mathcal{F}$ measurable, then $f^{-1}(\mathcal{E}) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}$. Conversely if $f^{-1}(\mathcal{E}) \subset \mathcal{M}$ then $\sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}$ and so making use of Lemma 6.3,

$$f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}.$$

■

Corollary 6.11. *Suppose that $(X,\mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *f is $(\mathcal{M},\mathcal{B}_{\mathbb{R}})$ – measurable,*
2. *$f^{-1}((a,\infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a,\infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty,a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Exercise 6.2. Prove Corollary 6.11. **Hint:** See Exercise 3.7.

Exercise 6.3. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 6.4. Let $(X,\mathcal{M})$ be a measure space and $f_n : X \to \mathbb{R}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{R}\} \in \mathcal{M}$. Similarly show the same holds if $\mathbb{R}$ is replaced by $\mathbb{C}$.

Exercise 6.5. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_{\mathbb{R}},\mathcal{B}_{\mathbb{R}})$ – measurable.

Definition 6.12. *Given measurable spaces $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff f is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

Proposition 6.13 (Localizing Measurability). *Let $(X,\mathcal{M})$ and $(Y,\mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If f is measurable and $A \subset X$ then $f|_A : A \to Y$ is $\mathcal{M}_A/\mathcal{F}$ – measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}/\mathcal{F}$ – measurable for all n, then f is $\mathcal{M}$ – measurable.*

Proof. 1. If $f : X \to Y$ is measurable, $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{F}$ and therefore

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathcal{M}_A \text{ for all } B \in \mathcal{F}.$$

2. If $B \in \mathcal{F}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty \left(f^{-1}(B) \cap A_n\right) = \cup_{n=1}^\infty f|_{A_n}^{-1}(B).$$

Since each $A_n \in \mathcal{M}$, $\mathcal{M}_{A_n} \subset \mathcal{M}$ and so the previous displayed equation shows $f^{-1}(B) \in \mathcal{M}$. ■

Lemma 6.14 (Composing Measurable Functions). *Suppose that $(X,\mathcal{M})$, $(Y,\mathcal{F})$ and $(Z,\mathcal{G})$ are measurable spaces. If $f : (X,\mathcal{M}) \to (Y,\mathcal{F})$ and $g : (Y,\mathcal{F}) \to (Z,\mathcal{G})$ are measurable functions then $g \circ f : (X,\mathcal{M}) \to (Z,\mathcal{G})$ is measurable as well.*

Proof. By assumption $g^{-1}(\mathcal{G}) \subset \mathcal{F}$ and $f^{-1}(\mathcal{F}) \subset \mathcal{M}$ so that

$$(g \circ f)^{-1}(\mathcal{G}) = f^{-1}\left(g^{-1}(\mathcal{G})\right) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}.$$

■

Definition 6.15 (σ – Algebras Generated by Functions). *Let X be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha,\mathcal{F}_\alpha) : \alpha \in I\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$. Let $\sigma(f_\alpha : \alpha \in I)$ denote the smallest σ – algebra on X such that each f_α is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

Example 6.16. Suppose that Y is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1,\ldots,y_N) = y_i$. Then, as the reader should check,

$$\sigma(\pi_1,\ldots,\pi_n) = \left\{A \times \Lambda^{N-n} : A \subset \Lambda^n\right\}.$$

Proposition 6.17. *Assuming the notation in Definition 6.15 (so $f_\alpha : X \to Y_\alpha$ for all $\alpha \in I$) and additionally let $(Z,\mathcal{M})$ be a measurable space. Then $g : Z \to X$ is $(\mathcal{M},\sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ g$ $\left(Z \xrightarrow{g} X \xrightarrow{f_\alpha} Y_\alpha\right)$ is $(\mathcal{M},\mathcal{F}_\alpha)$–measurable for all $\alpha \in I$.*

Proof. $(\Rightarrow)$ If g is $(\mathcal{M},\sigma(f_\alpha : \alpha \in I))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M},\mathcal{F}_\alpha)$ – measurable by Lemma 6.14.

(⇐) Since $\sigma(f_\alpha : \alpha \in I) = \sigma(\mathcal{E})$ where $\mathcal{E} := \cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)$, according to Proposition 6.10, it suffices to show $g^{-1}(A) \in \mathcal{M}$ for $A \in f_\alpha^{-1}(\mathcal{F}_\alpha)$. But this is true since if $A = f_\alpha^{-1}(B)$ for some $B \in \mathcal{F}_\alpha$, then $g^{-1}(A) = g^{-1}\left(f_\alpha^{-1}(B)\right) = (f_\alpha \circ g)^{-1}(B) \in \mathcal{M}$ because $f_\alpha \circ g : Z \to Y_\alpha$ is assumed to be measurable. ■

Definition 6.18. *If $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, then the product measure space, $(Y, \mathcal{F})$, is $Y := \prod_{\alpha \in I} Y_\alpha$, $\mathcal{F} := \sigma(\pi_\alpha : \alpha \in I)$ where $\pi_\alpha : Y \to Y_\alpha$ is the α – component projection. We call $\mathcal{F}$ the product σ – algebra and denote it by, $\mathcal{F} = \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

If A is a finite or countable set it is easily seen that

$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \sigma\left(\left\{\prod_{\alpha \in I} B_\alpha : B_\alpha \in \mathcal{F}_\alpha \text{ for all } \alpha \in I\right\}\right).$$

Let us record an important special case of Proposition 6.17.

Corollary 6.19. *If $(Z, \mathcal{M})$ is a measure space, then $g : Z \to Y = \prod_{\alpha \in I} Y_\alpha$ is $(\mathcal{M}, \mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha)$ – measurable iff $\pi_\alpha \circ g : Z \to Y_\alpha$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$.*

As a special case of the above corollary, if $A = \{1, 2, \ldots, n\}$, then $Y = Y_1 \times \cdots \times Y_n$ and $g = (g_1, \ldots, g_n) : Z \to Y$ is measurable iff each component, $g_i : Z \to Y_i$, is measurable. Here is another closely related result.

Proposition 6.20. *Suppose X is a set, $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in I\}$ is a collection of measurable spaces, and we are given maps, $f_\alpha : X \to Y_\alpha$, for all $\alpha \in I$. If $f : X \to Y := \prod_{\alpha \in I} Y_\alpha$ is the unique map, such that $\pi_\alpha \circ f = f_\alpha$, then*

$$\sigma(f_\alpha : \alpha \in I) = \sigma(f) = f^{-1}(\mathcal{F})$$

where $\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha$.

Proof. Since $\pi_\alpha \circ f = f_\alpha$ is $\sigma(f_\alpha : \alpha \in I)/\mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ it follows from Corollary 6.19 that $f : X \to Y$ is $\sigma(f_\alpha : \alpha \in I)/\mathcal{F}$ – measurable. Since $\sigma(f)$ is the smallest σ – algebra on X such that f is measurable we may conclude that $\sigma(f) \subset \sigma(f_\alpha : \alpha \in I)$.

Conversely, for each $\alpha \in I$, $f_\alpha = \pi_\alpha \circ f$ is $\sigma(f)/\mathcal{F}_\alpha$ – measurable for all $\alpha \in I$ being the composition of two measurable functions. Since $\sigma(f_\alpha : \alpha \in I)$ is the smallest σ – algebra on X such that each $f_\alpha : X \to Y_\alpha$ is measurable, we learn that $\sigma(f_\alpha : \alpha \in I) \subset \sigma(f)$. ■

Exercise 6.6. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\mathcal{E}_i$ is a subset of $\mathcal{F}_i$ such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for $i = 1$ and 2. Show $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{E})$ where $\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{E}_i \text{ for } i = 1, 2\}$. **Hints:**

1. First show that if Y is a set and $\mathcal{S}_1$ and $\mathcal{S}_2$ are two non-empty subsets of 2^Y, then $\sigma(\sigma(\mathcal{S}_1) \cup \sigma(\mathcal{S}_2)) = \sigma(\mathcal{S}_1 \cup \mathcal{S}_2)$. (In fact, one has that $\sigma(\cup_{\alpha \in I} \sigma(\mathcal{S}_\alpha)) = \sigma(\cup_{\alpha \in I} \mathcal{S}_\alpha)$ for any collection of non-empty subsets, $\{\mathcal{S}_\alpha\}_{\alpha \in I} \subset 2^Y$.)
2. After this you might start your proof as follows;

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\left(\pi_1^{-1}(\mathcal{F}_1) \cup \pi_2^{-1}(\mathcal{F}_2)\right) = \sigma\left(\pi_1^{-1}(\sigma(\mathcal{E}_2)) \cup \pi_2^{-1}(\sigma(\mathcal{E}_2))\right) = \ldots.$$

Remark 6.21. The reader should convince herself that Exercise 6.6 admits the following extension. If I is any finite or countable index set, $\{(Y_i, \mathcal{F}_i)\}_{i \in I}$ are measurable spaces and $\mathcal{E}_i \subset \mathcal{F}_i$ are such that $Y_i \in \mathcal{E}_i$ and $\mathcal{F}_i = \sigma(\mathcal{E}_i)$ for all $i \in I$, then

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j \text{ for all } j \in I\right\}\right)$$

Exercise 6.7. Suppose that $(Y_1, \mathcal{F}_1)$ and $(Y_2, \mathcal{F}_2)$ are measurable spaces and $\emptyset \neq B_i \subset Y_i$ for $i = 1, 2$. Show

$$[\mathcal{F}_1 \otimes \mathcal{F}_2]_{B_1 \times B_2} = [\mathcal{F}_1]_{B_1} \otimes [\mathcal{F}_2]_{B_2}.$$

Hint: you may find it useful to use the result of Exercise 6.6 with

$$\mathcal{E} := \{A_1 \times A_2 : A_i \in \mathcal{F}_i \text{ for } i = 1, 2\}.$$

Definition 6.22. *A function $f : X \to Y$ between two topological spaces is* ***Borel measurable*** *if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 6.23. *Let X and Y be two topological spaces and $f : X \to Y$ be a continuous function. Then f is Borel measurable.*

Proof. Using Lemma 6.3 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

■

Example 6.24. For $i = 1, 2, \ldots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each π_i is continuous and therefore $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable.

Lemma 6.25. *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \ldots, \pi_n) = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ and in particular, $A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$. (Also see Remark 6.21.)*

Proof. Assertion 1. Since $\mathcal{E} \subset \mathcal{B}_{\mathbb{R}^n}$, it follows that $\sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}$. Let

$$\mathcal{E}_0 := \{(a,b) : a, b \in \mathbb{Q}^n \ni a < b\},$$

where, for $a, b \in \mathbb{R}^n$, we write $a < b$ iff $a_i < b_i$ for $i = 1, 2, \dots, n$ and let

$$(a,b) = (a_1, b_1) \times \dots \times (a_n, b_n). \tag{6.3}$$

Since every open set, $V \subset \mathbb{R}^n$, may be written as a (necessarily) countable union of elements from $\mathcal{E}_0$, we have

$$V \in \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}),$$

i.e. $\sigma(\mathcal{E}_0)$ and hence $\sigma(\mathcal{E})$ contains all open subsets of $\mathbb{R}^n$. Hence we may conclude that

$$\mathcal{B}_{\mathbb{R}^n} = \sigma(\text{open sets}) \subset \sigma(\mathcal{E}_0) \subset \sigma(\mathcal{E}) \subset \mathcal{B}_{\mathbb{R}^n}.$$

Assertion 2. Since each $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is continuous, it is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable and therefore, $\sigma(\pi_1, \dots, \pi_n) \subset \mathcal{B}_{\mathbb{R}^n}$. Moreover, if (a,b) is as in Eq. (6.3), then

$$(a,b) = \cap_{i=1}^n \pi_i^{-1}((a_i, b_i)) \in \sigma(\pi_1, \dots, \pi_n).$$

Therefore, $\mathcal{E} \subset \sigma(\pi_1, \dots, \pi_n)$ and $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E}) \subset \sigma(\pi_1, \dots, \pi_n)$.

Assertion 3. If $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \dots, n$, then

$$A_1 \times \dots \times A_n = \cap_{i=1}^n \pi_i^{-1}(A_i) \in \sigma(\pi_1, \dots, \pi_n) = \mathcal{B}_{\mathbb{R}^n}.$$

■

Corollary 6.26. *If $(X, \mathcal{M})$ is a measurable space, then*

$$f = (f_1, f_2, \dots, f_n) : X \to \mathbb{R}^n$$

is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}^n})$ – measurable iff $f_i : X \to \mathbb{R}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable for each i. In particular, a function $f : X \to \mathbb{C}$ is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Proof. This is an application of Lemma 6.25 and Corollary 6.19 with $Y_i = \mathbb{R}$ for each i. ■

Corollary 6.27. *Let $(X, \mathcal{M})$ be a measurable space and $f, g : X \to \mathbb{C}$ be $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_\pm : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x), g(x))$, $A_\pm(w,z) = w \pm z$ and $M(w,z) = wz$. Then $A_\pm$ and M are continuous and hence $(\mathcal{B}_{\mathbb{C}^2}, \mathcal{B}_{\mathbb{C}})$ – measurable. Also F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_\pm \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ■

Lemma 6.28. *Let $\alpha \in \mathbb{C}$, $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{C}$ be a $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable function. Then*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if} \quad f(x) \neq 0 \\ \alpha & \text{if} \quad f(x) = 0 \end{cases}$$

is measurable.

Proof. Define $i : \mathbb{C} \to \mathbb{C}$ by

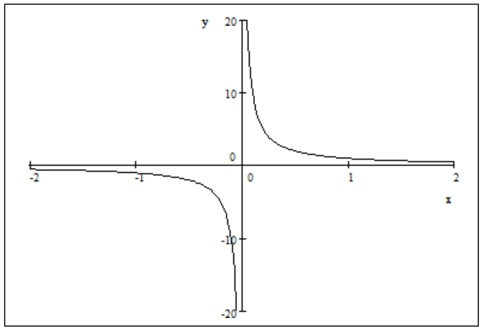
$$i(z) = \begin{cases} \frac{1}{z} & \text{if} \quad z \neq 0 \\ 0 & \text{if} \quad z = 0. \end{cases}$$

For any open set $V \subset \mathbb{C}$ we have

$$i^{-1}(V) = i^{-1}(V \setminus \{0\}) \cup i^{-1}(V \cap \{0\})$$

Because i is continuous except at $z = 0$, $i^{-1}(V \setminus \{0\})$ is an open set and hence in $\mathcal{B}_{\mathbb{C}}$. Moreover, $i^{-1}(V \cap \{0\}) \in \mathcal{B}_{\mathbb{C}}$ since $i^{-1}(V \cap \{0\})$ is either the empty set or the one point set $\{0\}$. Therefore $i^{-1}(\tau_{\mathbb{C}}) \subset \mathcal{B}_{\mathbb{C}}$ and hence $i^{-1}(\mathcal{B}_{\mathbb{C}}) = i^{-1}(\sigma(\tau_{\mathbb{C}})) = \sigma(i^{-1}(\tau_{\mathbb{C}})) \subset \mathcal{B}_{\mathbb{C}}$ which shows that i is Borel measurable. Since $F = i \circ f$ is the composition of measurable functions, F is also measurable. ■

Remark 6.29. For the real case of Lemma 6.28, define i as above but now take z to real. From the plot of i, Figure 6.29, the reader may easily verify that $i^{-1}((-\infty, a])$ is an infinite half interval for all a and therefore i is measurable. See Example 6.34 for another proof of this fact.



We will often deal with functions $f : X \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. When talking about measurability in this context we will refer to the σ – algebra on $\bar{\mathbb{R}}$ defined by

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma\left(\{[a,\infty] : a \in \mathbb{R}\}\right). \tag{6.4}$$

Proposition 6.30 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{6.5}$$

In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.

Proof. Let us first observe that

$$\{-\infty\} = \cap_{n=1}^{\infty}[-\infty,-n) = \cap_{n=1}^{\infty}[-n,\infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}},$$
$$\{\infty\} = \cap_{n=1}^{\infty}[n,\infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}.$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

$$\begin{aligned} i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) &= \sigma\left(i^{-1}\left(\left\{[a,\infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}\left([a,\infty]\right) : a \in \bar{\mathbb{R}}\right\}\right) \\ &= \sigma\left(\left\{[a,\infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a,\infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}. \end{aligned}$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}\left(\mathcal{B}_{\bar{\mathbb{R}}}\right) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \Longrightarrow A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (6.5). ∎

The proofs of the next two corollaries are left to the reader, see Exercises 6.8 and 6.9.

Corollary 6.31. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \bar{\mathbb{R}}$ be a function. Then the following are equivalent*

1. *f is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ - measurable,*
2. *$f^{-1}((a,\infty]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((-\infty,a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
4. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^0 : X \to \mathbb{R}$ defined by*

$$f^0(x) := \begin{cases} f(x) & \text{if } \quad f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm\infty\} \end{cases}$$

is measurable.

Corollary 6.32. *Let $(X, \mathcal{M})$ be a measurable space, $f, g : X \to \bar{\mathbb{R}}$ be functions and define $f \cdot g : X \to \bar{\mathbb{R}}$ and $(f+g) : X \to \bar{\mathbb{R}}$ using the conventions, $0 \cdot \infty = 0$ and $(f+g)(x) = 0$ if $f(x) = \infty$ and $g(x) = -\infty$ or $f(x) = -\infty$ and $g(x) = \infty$. Then $f \cdot g$ and $f + g$ are measurable functions on X if both f and g are measurable.*

Exercise 6.8. Prove Corollary 6.31 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 6.11. Use Proposition 6.30 to handle item 4.

Exercise 6.9. Prove Corollary 6.32.

Proposition 6.33 (Closure under sups, infs and limits). *Suppose that $(X, \mathcal{M})$ is a measurable space and $f_j : (X, \mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*

$$\sup{}_j f_j, \quad \inf{}_j f_j, \quad \limsup_{j\to\infty} \ f_j \text{ and } \liminf_{j\to\infty} \ f_j$$

are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X, \mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)

Proof. Define $g_+(x) := \sup{}_j \ f_j(x)$, then

$$\begin{aligned} \{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \ \forall j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M} \end{aligned}$$

so that g_+ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then

$$\{x : g_-(x) \ge a\} = \cap_j \{x : f_j(x) \ge a\} \in \mathcal{M}.$$

Since

$$\limsup_{j\to\infty} \ f_j = \inf_n \sup\{f_j : j \ge n\} \text{ and}$$
$$\liminf_{j\to\infty} \ f_j = \sup_n \inf\{f_j : j \ge n\}$$

we are done by what we have already proved. ∎

Example 6.34. As we saw in Remark 6.29, $i : \mathbb{R} \to \mathbb{R}$ defined by

$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

is measurable by a simple direct argument. For an alternative argument, let

$$i_n(z) := \frac{z}{z^2 + \frac{1}{n}} \text{ for all } n \in \mathbb{N}.$$

Then i_n is continuous and $\lim_{n\to\infty} i_n(z) = i(z)$ for all $z \in \mathbb{R}$ from which it follows that i is Borel measurable.

Example 6.35. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q}\cap[0,1]$ and define

$$f(x)=\sum_{n=1}^\infty 2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x-r_n|}}=5 \text{ if } x=r_n.$$

Then $f:\mathbb{R}\to\bar{\mathbb{R}}$ is measurable. Indeed, if

$$g_n(x)=\begin{cases}\frac{1}{\sqrt{|x-r_n|}} & \text{if } x\neq r_n\\ 0 & \text{if } x=r_n\end{cases}$$

then $g_n(x)=\sqrt{|i(x-r_n)|}$ is measurable as the composition of measurable is measurable. Therefore $g_n+5\cdot 1_{\{r_n\}}$ is measurable as well. Finally,

$$f(x)=\lim_{N\to\infty}\sum_{n=1}^N 2^{-n}\frac{1}{\sqrt{|x-r_n|}}$$

is measurable since sums of measurable functions are measurable and limits of measurable functions are measurable. **Moral:** if you can explicitly write a function $f:\bar{\mathbb{R}}\to\bar{\mathbb{R}}$ down then it is going to be measurable.

Definition 6.36. *Given a function* $f:X\to\bar{\mathbb{R}}$ *let* $f_+(x):=\max\{f(x),0\}$ *and* $f_-(x):=\max(-f(x),0)=-\min(f(x),0)$. *Notice that* $f=f_+-f_-$.

Corollary 6.37. *Suppose* $(X,\mathcal{M})$ *is a measurable space and* $f:X\to\bar{\mathbb{R}}$ *is a function. Then* f *is measurable iff* $f_\pm$ *are measurable.*

Proof. If f is measurable, then Proposition 6.33 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f=f_+-f_-$. ■

Definition 6.38. *Let* $(X,\mathcal{M})$ *be a measurable space. A function* $\varphi:X\to\mathbb{F}$ *(*$\mathbb{F}$ *denotes either* $\mathbb{R}$, $\mathbb{C}$ *or* $[0,\infty]\subset\bar{\mathbb{R}}$*) is a* ***simple function*** *if* φ *is* $\mathcal{M}-\mathcal{B}_\mathbb{F}$ *measurable and* $\varphi(X)$ *contains only finitely many elements.*

Any such simple functions can be written as

$$\varphi=\sum_{i=1}^n\lambda_i 1_{A_i} \text{ with } A_i\in\mathcal{M} \text{ and } \lambda_i\in\mathbb{F}. \tag{6.6}$$

Indeed, take $\lambda_1,\lambda_2,\dots,\lambda_n$ to be an enumeration of the range of φ and $A_i=\varphi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\varphi=\sum_{y\in\mathbb{F}}y1_{\varphi^{-1}(\{y\})}. \tag{6.7}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

Theorem 6.39 (Approximation Theorem). *Let* $f:X\to[0,\infty]$ *be measurable and define, see Figure 6.1,*

$$\begin{aligned}\varphi_n(x)&:=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{f^{-1}\left(\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]\right)}(x)+2^n1_{f^{-1}((2^n,\infty])}(x)\\&=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\left\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right\}}(x)+2^n1_{\{f>2^n\}}(x)\end{aligned}$$

then $\varphi_n\le f$ *for all* n, $\varphi_n(x)\uparrow f(x)$ *for all* $x\in X$ *and* $\varphi_n\uparrow f$ *uniformly on the sets* $X_M:=\{x\in X:f(x)\le M\}$ *with* $M<\infty$.

Moreover, if $f:X\to\mathbb{C}$ *is a measurable function, then there exists simple functions* φ_n *such that* $\lim_{n\to\infty}\varphi_n(x)=f(x)$ *for all* x *and* $|\varphi_n|\uparrow|f|$ *as* $n\to\infty$.

Proof. Since $f^{-1}\left(\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]\right)$ and $f^{-1}((2^n,\infty])$ are in $\mathcal{M}$ as f is measurable, φ_n is a measurable simple function for each n. Because

$$\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]=\left(\frac{2k}{2^{n+1}},\frac{2k+1}{2^{n+1}}\right]\cup\left(\frac{2k+1}{2^{n+1}},\frac{2k+2}{2^{n+1}}\right],$$

if $x\in f^{-1}\left(\left(\frac{2k}{2^{n+1}},\frac{2k+1}{2^{n+1}}\right]\right)$ then $\varphi_n(x)=\varphi_{n+1}(x)=\frac{2k}{2^{n+1}}$ and if $x\in f^{-1}\left(\left(\frac{2k+1}{2^{n+1}},\frac{2k+2}{2^{n+1}}\right]\right)$ then $\varphi_n(x)=\frac{2k}{2^{n+1}}<\frac{2k+1}{2^{n+1}}=\varphi_{n+1}(x)$. Similarly

$$(2^n,\infty]=(2^n,2^{n+1}]\cup(2^{n+1},\infty],$$

and so for $x\in f^{-1}((2^{n+1},\infty])$, $\varphi_n(x)=2^n<2^{n+1}=\varphi_{n+1}(x)$ and for $x\in f^{-1}((2^n,2^{n+1}])$, $\varphi_{n+1}(x)\ge 2^n=\varphi_n(x)$. Therefore $\varphi_n\le\varphi_{n+1}$ for all n. It is clear by construction that $0\le\varphi_n(x)\le f(x)$ for all x and that $0\le f(x)-\varphi_n(x)\le 2^{-n}$ if $x\in X_{2^n}=\{f\le 2^n\}$. Hence we have shown that $\varphi_n(x)\uparrow f(x)$ for all $x\in X$ and $\varphi_n\uparrow f$ uniformly on bounded sets.

For the second assertion, first assume that $f:X\to\mathbb{R}$ is a measurable function and choose $\varphi_n^\pm$ to be non-negative simple functions such that $\varphi_n^\pm\uparrow f_\pm$ as $n\to\infty$ and define $\varphi_n=\varphi_n^+-\varphi_n^-$. Then (using $\varphi_n^+\cdot\varphi_n^-\le f_+\cdot f_-=0$)

$$|\varphi_n|=\varphi_n^++\varphi_n^-\le\varphi_{n+1}^++\varphi_{n+1}^-=|\varphi_{n+1}|$$

and clearly $|\varphi_n|=\varphi_n^++\varphi_n^-\uparrow f_++f_-=|f|$ and $\varphi_n=\varphi_n^+-\varphi_n^-\to f_+-f_-=f$ as $n\to\infty$. Now suppose that $f:X\to\mathbb{C}$ is measurable. We may now choose simple function u_n and v_n such that $|u_n|\uparrow|\mathrm{Re} f|$, $|v_n|\uparrow|\mathrm{Im} f|$, $u_n\to\mathrm{Re} f$ and $v_n\to\mathrm{Im} f$ as $n\to\infty$. Let $\varphi_n=u_n+iv_n$, then

$$|\varphi_n|^2=u_n^2+v_n^2\uparrow|\mathrm{Re} f|^2+|\mathrm{Im} f|^2=|f|^2$$

and $\varphi_n=u_n+iv_n\to\mathrm{Re} f+i\,\mathrm{Im} f=f$ as $n\to\infty$. ■

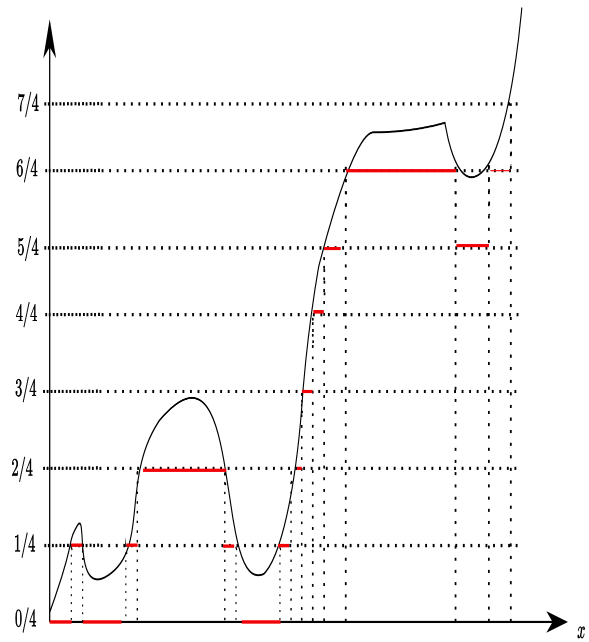
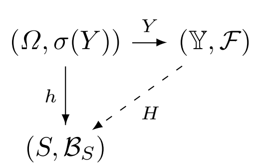


Fig. 6.1. Constructing the simple function, φ_2, approximating a function, $f : X \to [0,\infty]$. The graph of φ_2 is in red.

6.2 Factoring Random Variables

Lemma 6.40. *Suppose that $(\mathbb{Y},\mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y),\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $h : \Omega \to \bar{\mathbb{R}}$, there is a $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $H : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $h = H \circ Y$. More generally, $\bar{\mathbb{R}}$ may be replaced by any "standard Borel space,"*[1] *i.e. a space, $(S,\mathcal{B}_S)$ which is measure theoretic isomorphic to a Borel subset of $\mathbb{R}$.*



Proof. First suppose that $h = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $H = 1_B$. More generally if $h = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $h = H \circ Y$ with $H := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, h, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions h_n converging to h. Let $H_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $h_n = H_n \circ Y$. Then it follows that

$$h = \lim_{n\to\infty} h_n = \limsup_{n\to\infty} h_n = \limsup_{n\to\infty} H_n \circ Y = H \circ Y$$

where $H := \limsup_{n\to\infty} H_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$.

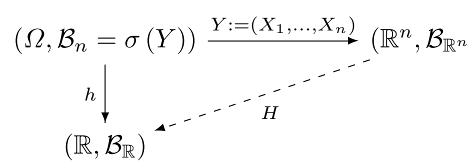
For the last assertion we may assume that $S \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_S = (\mathcal{B}_{\mathbb{R}})_S = \{A \cap S : A \in \mathcal{B}_{\mathbb{R}}\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \bar{\mathbb{R}}$ which is $(\mathcal{F},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable such that $h = i_S \circ h = H \circ Y$. The only problems with H is that $H(\mathbb{Y})$ may not be contained in S. To fix this, let

$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on } \quad H^{-1}(S) \\ \quad * & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

where $*$ is some fixed arbitrary point in S. It follows from Proposition 6.13 that $H_S : \mathbb{Y} \to S$ is $(\mathcal{F},\mathcal{B}_S)$ – measurable and we still have $h = H_S \circ Y$ as the range of Y must necessarily be in $H^{-1}(S)$. ∎

Here is how this lemma will often be used in these notes.

Corollary 6.41. *Suppose that $(\Omega,\mathcal{B})$ is a measurable space, $X_n : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ – measurable functions, and $\mathcal{B}_n := \sigma(X_1,\dots,X_n) \subset \mathcal{B}$ for each $n \in \mathbb{N}$. Then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff there exists $H : \mathbb{R}^n \to \mathbb{R}$ which is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable such that $h = H(X_1,\dots,X_n)$.*



Proof. By Lemma 6.25 and Corollary 6.19, the map, $Y := (X_1,\dots,X_n) : \Omega \to \mathbb{R}^n$ is $(\mathcal{B},\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}})$ – measurable and by Proposition 6.20, $\mathcal{B}_n = \sigma(X_1,\dots,X_n) = \sigma(Y)$. Thus we may apply Lemma 6.40 to see that there exists a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$, such that $h = H \circ Y = H(X_1,\dots,X_n)$. ∎

[1] Standard Borel spaces include almost any measurable space that we will consider in these notes. For example they include all complete seperable metric spaces equipped with the Borel σ – algebra, see Section **??**.

6.3 Summary of Measurability Statements

It may be worthwhile to gather the statements of the main measurability results of Sections 6.1 and 6.2 in one place. To do this let $(\Omega, \mathcal{B})$, $(X, \mathcal{M})$, and $\{(Y_\alpha, \mathcal{F}_\alpha)\}_{\alpha \in I}$ be measurable spaces and $f_\alpha : \Omega \to Y_\alpha$ be given maps for all $\alpha \in I$. Also let $\pi_\alpha : Y \to Y_\alpha$ be the α – projection map,

$$\mathcal{F} := \otimes_{\alpha \in I} \mathcal{F}_\alpha := \sigma(\pi_\alpha : \alpha \in I)$$

be the product σ – algebra on Y, and $f : \Omega \to Y$ be the unique map determined by $\pi_\alpha \circ f = f_\alpha$ for all $\alpha \in I$. Then the following measurability results hold;

1. For $A \subset \Omega$, the indicator function, 1_A, is $(\mathcal{B}, \mathcal{B}_{\mathbb{R}})$ – measurable iff $A \in \mathcal{B}$. (Example 6.8).
2. If $\mathcal{E} \subset \mathcal{M}$ generates $\mathcal{M}$ (i.e. $\mathcal{M} = \sigma(\mathcal{E})$), then a map, $g : \Omega \to X$ is $(\mathcal{B}, \mathcal{M})$ – measurable iff $g^{-1}(\mathcal{E}) \subset \mathcal{B}$ (Lemma 6.3 and Proposition 6.10).
3. The notion of measurability may be localized (Proposition 6.13).
4. Composition of measurable functions are measurable (Lemma 6.14).
5. Continuous functions between two topological spaces are also Borel measurable (Proposition 6.23).
6. $\sigma(f) = \sigma(f_\alpha : \alpha \in I)$ (Proposition 6.20).
7. A map, $h : X \to \Omega$ is $(\mathcal{M}, \sigma(f) = \sigma(f_\alpha : \alpha \in I))$ – measurable iff $f_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Proposition 6.17).
8. A map, $h : X \to Y$ is $(\mathcal{M}, \mathcal{F})$ – measurable iff $\pi_\alpha \circ h$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all $\alpha \in I$ (Corollary 6.19).
9. If $I = \{1, 2, \dots, n\}$, then
$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma(\{A_1 \times A_2 \times \cdots \times A_n : A_i \in \mathcal{F}_i \text{ for } i \in I\}),$$
this is a special case of Remark 6.21.
10. $\mathcal{B}_{\mathbb{R}^n} = \mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}}$ (n - times) for all $n \in \mathbb{N}$, i.e. the Borel σ – algebra on $\mathbb{R}^n$ is the same as the product σ – algebra. (Lemma 6.25).
11. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ is closed under the usual pointwise algebraic operations (Corollary 6.32). They are also closed under the countable supremums, infimums, and limits (Proposition 6.33).
12. The collection of measurable functions from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ is closed under the usual pointwise algebraic operations and countable limits. (Corollary 6.27 and Proposition 6.33). The limiting assertion follows by considering the real and imaginary parts of all functions involved.
13. The class of measurable functions from $(\Omega, \mathcal{B})$ to $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ and from $(\Omega, \mathcal{B})$ to $(\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ may be well approximated by measurable simple functions (Theorem 6.39).
14. If $X_i : \Omega \to \mathbb{R}$ are $\mathcal{B}/\mathcal{B}_{\mathbb{R}}$ – measurable maps and $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff $h = H(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable map, $H : \mathbb{R}^n \to \mathbb{R}$ (Corollary 6.41).
15. We also have the more general factorization Lemma 6.40.

For the most part most of our future measurability issues can be resolved by one or more of the items on this list.

6.4 Distributions / Laws of Random Vectors

The proof of the following proposition is routine and will be left to the reader.

Proposition 6.42. *Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$. Then ν is a measure on $(Y, \mathcal{F})$. (In the future we will denote ν by $f_*\mu$ or $\mu \circ f^{-1}$ and call $f_*\mu$ the* ***push-forward of*** *μ* ***by*** *f or the* ***law of*** *f* ***under*** *μ.*

Definition 6.43. *Suppose that $\{X_i\}_{i=1}^n$ is a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$. The probability measure,*

$$\mu = (X_1, \dots, X_n)_* P = P \circ (X_1, \dots, X_n)^{-1} \text{ on } \mathcal{B}_{\mathbb{R}}$$

(see Proposition 6.42) is called the ***joint distribution*** *(or* ***law****) of $(X_1, \dots, X_n)$. To be more explicit,*

$$\mu(B) := P((X_1, \dots, X_n) \in B) := P(\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in B\})$$

for all $B \in \mathcal{B}_{\mathbb{R}^n}$.

Corollary 6.44. *The joint distribution, μ is uniquely determined from the knowledge of*

$$P((X_1, \dots, X_n) \in A_1 \times \cdots \times A_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

or from the knowledge of

$$P(X_1 \le x_1, \dots, X_n \le x_n) \text{ for all } A_i \in \mathcal{B}_{\mathbb{R}}$$

for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Proof. Apply Proposition 5.15 with $\mathcal{P}$ being the π – systems defined by

$$\mathcal{P} := \{A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$$

for the first case and

$$\mathcal{P} := \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \in \mathcal{B}_{\mathbb{R}^n} : x_i \in \mathbb{R}\}$$

for the second case. ∎

Definition 6.45. *Suppose that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two finite sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ respectively. We write $(X_1, \dots, X_n) \stackrel{d}{=} (Y_1, \dots, Y_n)$ if $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the* ***same distribution****, i.e. if*

$$P((X_1, \dots, X_n) \in B) = P'((Y_1, \dots, Y_n) \in B) \text{ for all } B \in \mathcal{B}_{\mathbb{R}^n}.$$

More generally, if $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ are two sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ we write $\{X_i\}_{i=1}^\infty \stackrel{d}{=} \{Y_i\}_{i=1}^\infty$ iff $(X_1, \dots, X_n) \stackrel{d}{=} (Y_1, \dots, Y_n)$ for all $n \in \mathbb{N}$.

Proposition 6.46. *Let us continue using the notation in Definition 6.45. Further let*

$$X = (X_1, X_2, \dots) : \Omega \to \mathbb{R}^{\mathbb{N}} \text{ and } Y := (Y_1, Y_2, \dots) : \Omega' \to \mathbb{R}^{\mathbb{N}}$$

and let $\mathcal{F} := \otimes_{n \in \mathbb{N}} \mathcal{B}_{\mathbb{R}}$ – be the product σ – algebra on $\mathbb{R}^{\mathbb{N}}$. Then $\{X_i\}_{i=1}^\infty \stackrel{d}{=} \{Y_i\}_{i=1}^\infty$ iff $X_*P = Y_*P'$ as measures on $(\mathbb{R}^{\mathbb{N}}, \mathcal{F})$.

Proof. Let

$$\mathcal{P} := \cup_{n=1}^\infty \left\{A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}} : A_i \in \mathcal{B}_{\mathbb{R}} \text{ for } 1 \le i \le n\right\}.$$

Notice that $\mathcal{P}$ is a π – system and it is easy to show $\sigma(\mathcal{P}) = \mathcal{F}$ (see Exercise 6.6). Therefore by Proposition 5.15, $X_*P = Y_*P'$ iff $X_*P = Y_*P'$ on $\mathcal{P}$. Now for $A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}} \in \mathcal{P}$ we have,

$$X_*P\left(A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R}^{\mathbb{N}}\right) = P((X_1, \dots, X_n) \in A_1 \times A_2 \times \cdots \times A_n)$$

and hence the condition becomes,

$$P((X_1, \dots, X_n) \in A_1 \times A_2 \times \cdots \times A_n) = P'((Y_1, \dots, Y_n) \in A_1 \times A_2 \times \cdots \times A_n)$$

for all $n \in \mathbb{N}$ and $A_i \in \mathcal{B}_{\mathbb{R}}$. Another application of Proposition 5.15 or using Corollary 6.44 allows us to conclude that shows that $X_*P = Y_*P'$ iff $(X_1, \dots, X_n) \stackrel{d}{=} (Y_1, \dots, Y_n)$ for all $n \in \mathbb{N}$. ∎

Corollary 6.47. *Continue the notation above and assume that $\{X_i\}_{i=1}^\infty \stackrel{d}{=} \{Y_i\}_{i=1}^\infty$. Further let*

$$X_\pm = \begin{cases} \limsup_{n\to\infty} X_n & \text{if } + \\ \liminf_{n\to\infty} X_n & \text{if } - \end{cases}$$

and define $Y_\pm$ similarly. Then $(X_-, X_+) \stackrel{d}{=} (Y_-, Y_+)$ as random variables into $\left(\bar{\mathbb{R}}^2, \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}\right)$. In particular,

$$P\left(\lim_{n\to\infty} X_n \text{ exists in } \mathbb{R}\right) = P'\left(\lim_{n\to\infty} Y \text{ exists in } \mathbb{R}\right). \tag{6.8}$$

Proof. First suppose that $(\Omega', \mathcal{B}', P') = \left(\mathbb{R}^{\mathbb{N}}, \mathcal{F}, P' := X_* P\right)$ where $Y_i(a_1, a_2, \dots) := a_i = \pi_i(a_1, a_2, \dots)$. Then for $C \in \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ we have,

$$X^{-1}(\{(Y_-, Y_+) \in C\}) = \{(Y_- \circ X, Y_+ \circ X) \in C\} = \{(X_-, X_+) \in C\},$$

since, for example,

$$Y_- \circ X = \liminf_{n\to\infty} Y_n \circ X = \liminf_{n\to\infty} X_n = X_-.$$

Therefore it follows that

$$P((X_-, X_+) \in C) = P \circ X^{-1}(\{(Y_-, Y_+) \in C\}) = P'(\{(Y_-, Y_+) \in C\}). \quad (6.9)$$

The general result now follows by two applications of this special case.

For the last assertion, take

$$C = \{(x, x) : x \in \mathbb{R}\} \in \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}} \otimes \mathcal{B}_{\bar{\mathbb{R}}}.$$

Then $(X_-, X_+) \in C$ iff $X_- = X_+ \in \mathbb{R}$ which happens iff $\lim_{n\to\infty} X_n$ exists in $\mathbb{R}$. Similarly, $(Y_-, Y_+) \in C$ iff $\lim_{n\to\infty} Y_n$ exists in $\mathbb{R}$ and therefore Eq. (6.8) holds as a consequence of Eq. (6.9). ■

Exercise 6.10. Let $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ be two sequences of random variables such that $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n := X_1 + \cdots + X_n$ and $T_n := Y_1 + \cdots + Y_n$. Prove the following assertions.

1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1, \dots, X_n) \overset{d}{=} f(Y_1, \dots, Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty \overset{d}{=} \{T_n\}_{n=1}^\infty$.
   Hint: Apply item 1. with $k = n$ after making a judicious choice for $f : \mathbb{R}^n \to \mathbb{R}^n$.

6.5 Generating All Distributions from the Uniform Distribution

Theorem 6.48. *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $G : (0,1) \to \mathbb{R}$ be defined (see Figure 6.2) by,*

$$G(y) := \inf\{x : F(x) \geq y\}.$$

Then $G : (0,1) \to \mathbb{R}$ is Borel measurable and $G_* m = \mu_F$ where μ_F is the unique measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

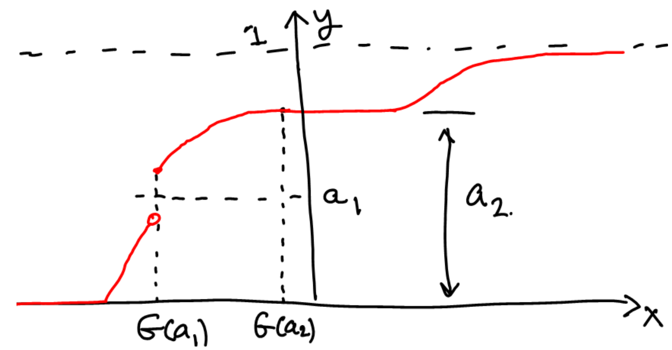


Fig. 6.2. A pictorial definition of G.

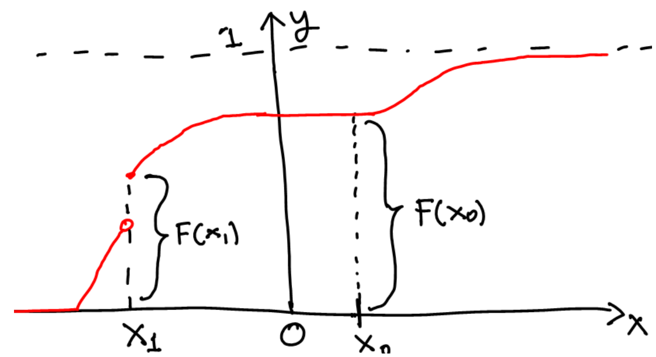


Fig. 6.3. As can be seen from this picture, $G(y) \leq x_0$ iff $y \leq F(x_0)$ and similarly, $G(y) \leq x_1$ iff $y \leq x_1$.

Proof. Since $G : (0,1) \to \mathbb{R}$ is a non-decreasing function, G is measurable. We also claim that, for all $x_0 \in \mathbb{R}$, that

$$G^{-1}((0, x_0]) = \{y : G(y) \leq x_0\} = (0, F(x_0)] \cap \mathbb{R}, \quad (6.10)$$

see Figure 6.3.

To give a formal proof of Eq. (6.10), $G(y) = \inf\{x : F(x) \geq y\} \leq x_0$, there exists $x_n \geq x_0$ with $x_n \downarrow x_0$ such that $F(x_n) \geq y$. By the right continuity of F, it follows that $F(x_0) \geq y$. Thus we have shown

$$\{G \leq x_0\} \subset (0, F(x_0)] \cap (0,1).$$

For the converse, if $y \leq F(x_0)$ then $G(y) = \inf\{x : F(x) \geq y\} \leq x_0$, i.e. $y \in \{G \leq x_0\}$. Indeed, $y \in G^{-1}((-\infty, x_0])$ iff $G(y) \leq x_0$. Observe that

$$G(F(x_0)) = \inf\{x : F(x) \geq F(x_0)\} \leq x_0$$

and hence $G(y) \leq x_0$ whenever $y \leq F(x_0)$. This shows that

$$(0, F(x_0)] \cap (0,1) \subset G^{-1}((0, x_0]).$$

As a consequence we have $G_* m = \mu_F$. Indeed,

$$\begin{aligned}(G_* m)((-\infty, x]) &= m\left(G^{-1}((-\infty, x])\right) = m(\{y \in (0,1) : G(y) \leq x\}) \\ &= m((0, F(x)] \cap (0,1)) = F(x).\end{aligned}$$

See section 2.5.2 on p. 61 of Resnick for more details. ■

Theorem 6.49 (Durret's Version). *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $Y : (0,1) \to \mathbb{R}$ be defined (see Figure 6.4) by,*

$$Y(x) := \sup\{y : F(y) < x\}.$$

Then $Y : (0,1) \to \mathbb{R}$ is Borel measurable and $Y_* m = \mu_F$ where μ_F is the unique measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

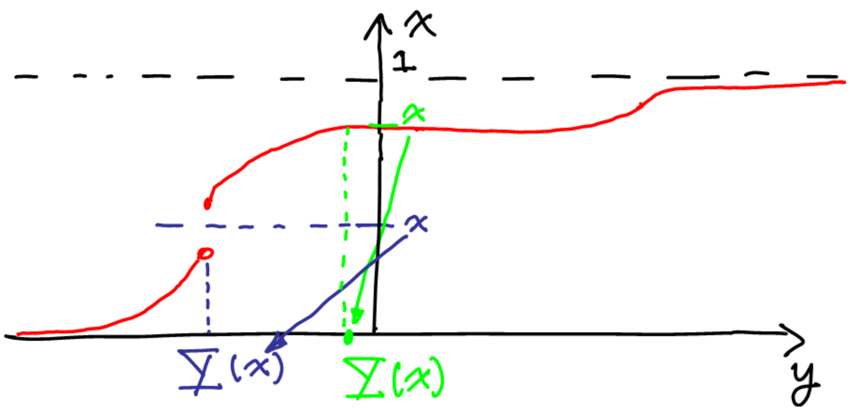


Fig. 6.4. A pictorial definition of $Y(x)$.

Proof. Since $Y : (0,1) \to \mathbb{R}$ is a non-decreasing function, Y is measurable. Also observe, if $y < Y(x)$, then $F(y) < x$ and hence,

$$F(Y(x)-) = \lim_{y \uparrow Y(x)} F(y) \leq x.$$

For $y > Y(x)$, we have $F(y) \geq x$ and therefore,

$$F(Y(x)) = F(Y(x)+) = \lim_{y \downarrow Y(x)} F(y) \geq x$$

and so we have shown

$$F(Y(x)-) \leq x \leq F(Y(x)).$$

We will now show

$$\{x \in (0,1) : Y(x) \leq y_0\} = (0, F(y_0)] \cap (0,1). \tag{6.11}$$

For the inclusion "$\subset$," if $x \in (0,1)$ and $Y(x) \leq y_0$, then $x \leq F(Y(x)) \leq F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely if $x \in (0,1)$ and $x \leq F(y_0)$ then (by definition of $Y(x)$) $y_0 \geq Y(x)$.

From the identity in Eq. (6.11), it follows that Y is measurable and

$$(Y_* m)((-\infty, y_0)) = m\left(Y^{-1}(-\infty, y_0)\right) = m((0, F(y_0)] \cap (0,1)) = F(y_0).$$

Therefore, $Law(Y) = \mu_F$ as desired. ■

7

Integration Theory

In this chapter, we will greatly extend the "simple" integral or expectation which was developed in Section 4.3 above. Recall there that if $(\Omega, \mathcal{B}, \mu)$ was measurable space and $f : \Omega \to [0, \infty]$ was a measurable simple function, then we let

$$\mathbb{E}_\mu f := \sum_{\lambda \in [0,\infty]} \lambda \mu (f = \lambda).$$

The conventions[1] being use here is that $\infty \cdot \mu (f = \infty) = 0$ if $\mu (f = \infty) = 0$ and $0 \cdot \mu (f = 0) = 0$ when $\mu (f = 0) = \infty$. In short, in these integration formulas we adopt the convention that $0 \cdot \infty = 0$. Please be careful not to apply this convention in general elsewhere.

7.1 Integrals of positive functions

Definition 7.1. *Let $L^+ = L^+ (\mathcal{B}) = \{f : \Omega \to [0, \infty] : f$ is measurable$\}$. Define*

$$\int_\Omega f (\omega) \, d\mu (\omega) = \int_\Omega f d\mu := \sup \{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\}.$$

We say the $f \in L^+$ is ***integrable*** *if $\int_\Omega f d\mu < \infty$. If $A \in \mathcal{B}$, let*

$$\int_A f (\omega) \, d\mu (\omega) = \int_A f d\mu := \int_\Omega 1_A f \; d\mu.$$

We also use the notation,

$$\mathbb{E} f = \int_\Omega f d\mu \text{ and } \mathbb{E} [f : A] := \int_A f d\mu.$$

Remark 7.2. Because of item 3. of Proposition 4.19, if φ is a non-negative simple function, $\int_\Omega \varphi d\mu = \mathbb{E}_\mu \varphi$ so that $\int_\Omega$ is an extension of $\mathbb{E}_\mu$.

Lemma 7.3. *Let $f, g \in L^+ (\mathcal{B})$. Then:*

1. *if $\lambda \ge 0$, then*
$$\int_\Omega \lambda f d\mu = \lambda \int_\Omega f d\mu$$
wherein $\lambda \int_\Omega f d\mu \equiv 0$ if $\lambda = 0$, even if $\int_\Omega f d\mu = \infty$.
2. *if $0 \le f \le g$, then*
$$\int_\Omega f d\mu \le \int_\Omega g d\mu. \tag{7.1}$$
3. *For all $\varepsilon > 0$ and $p > 0$,*
$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon^p} \int_\Omega f^p 1_{\{f \ge \varepsilon\}} d\mu \le \frac{1}{\varepsilon^p} \int_\Omega f^p d\mu. \tag{7.2}$$
The inequality in Eq. (7.2) is called Chebyshev's Inequality for $p = 1$ and Markov's inequality for $p = 2$.
4. *If $\int_\Omega f d\mu < \infty$ then $\mu(f = \infty) = 0$ (i.e. $f < \infty$ a.e.) and the set $\{f > 0\}$ is σ – finite.*

Proof. 1. We may assume $\lambda > 0$ in which case,

$$\begin{aligned}
\int_\Omega \lambda f d\mu &= \sup \{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le \lambda f\} \\
&= \sup \left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \lambda^{-1} \varphi \le f\right\} \\
&= \sup \{\mathbb{E}_\mu [\lambda \psi] : \psi \text{ is simple and } \psi \le f\} \\
&= \sup \{\lambda \mathbb{E}_\mu [\psi] : \psi \text{ is simple and } \psi \le f\} \\
&= \lambda \int_\Omega f d\mu.
\end{aligned}$$

2. Since

$$\{\varphi \text{ is simple and } \varphi \le f\} \subset \{\varphi \text{ is simple and } \varphi \le g\},$$

Eq. (7.1) follows from the definition of the integral.

3. Since $1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \frac{1}{\varepsilon} f \le \frac{1}{\varepsilon} f$ we have

$$1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \le \left(\frac{1}{\varepsilon} f\right)^p$$

[1] This is the convention necessary in order for the monotone convergence theorem to hold.

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \geq \varepsilon) = \int_\Omega 1_{\{f \geq \varepsilon\}} d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_\Omega 1_{\{f \geq \varepsilon\}} f^p d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_\Omega f^p d\mu.$$

4. If $\mu(f = \infty) > 0$, then $\varphi_n := n 1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \leq f$ for all n and hence

$$n\mu(f = \infty) = \mathbb{E}_\mu(\varphi_n) \leq \int_\Omega f d\mu$$

for all n. Letting $n \to \infty$ shows $\int_\Omega f d\mu = \infty$. Thus if $\int_\Omega f d\mu < \infty$ then $\mu(f = \infty) = 0$.

Moreover,

$$\{f > 0\} = \cup_{n=1}^\infty \{f > 1/n\}$$

with $\mu(f > 1/n) \leq n \int_\Omega f d\mu < \infty$ for each n. ■

Theorem 7.4 (Monotone Convergence Theorem). *Suppose $f_n \in L^+$ is a sequence of functions such that $f_n \uparrow f$ (f is necessarily in L^+) then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Since $f_n \leq f_m \leq f$, for all $n \leq m < \infty$,

$$\int f_n \leq \int f_m \leq \int f$$

from which if follows $\int f_n$ is increasing in n and

$$\lim_{n\to\infty} \int f_n \leq \int f. \tag{7.3}$$

For the opposite inequality, let $\varphi : \Omega \to [0, \infty)$ be a simple function such that $0 \leq \varphi \leq f$, $\alpha \in (0,1)$ and $\Omega_n := \{f_n \geq \alpha\varphi\}$. Notice that $\Omega_n \uparrow \Omega$ and $f_n \geq \alpha 1_{\Omega_n}\varphi$ and so by definition of $\int f_n$,

$$\int f_n \geq \mathbb{E}_\mu[\alpha 1_{\Omega_n}\varphi] = \alpha \mathbb{E}_\mu[1_{\Omega_n}\varphi]. \tag{7.4}$$

Then using the identity

$$1_{\Omega_n}\varphi = 1_{\Omega_n} \sum_{y>0} y 1_{\{\varphi=y\}} = \sum_{y>0} y 1_{\{\varphi=y\}\cap\Omega_n},$$

and the linearity of $\mathbb{E}_\mu$ we have,

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}_\mu[1_{\Omega_n}\varphi] &= \lim_{n\to\infty} \sum_{y>0} y \cdot \mu(\Omega_n \cap \{\varphi = y\}) \\ &= \sum_{y>0} y \lim_{n\to\infty} \mu(\Omega_n \cap \{\varphi = y\}) \text{ (finite sum)} \\ &= \sum_{y>0} y\mu(\{\varphi = y\}) = \mathbb{E}_\mu[\varphi], \end{aligned}$$

wherein we have used the continuity of μ under increasing unions for the third equality. This identity allows us to let $n \to \infty$ in Eq. (7.4) to conclude $\lim_{n\to\infty} \int f_n \geq \alpha \mathbb{E}_\mu[\varphi]$ and since $\alpha \in (0,1)$ was arbitrary we may further conclude, $\mathbb{E}_\mu[\varphi] \leq \lim_{n\to\infty} \int f_n$. The latter inequality being true for all simple functions φ with $\varphi \leq f$ then implies that

$$\int f = \sup_{0 \leq \varphi \leq f} \mathbb{E}_\mu[\varphi] \leq \lim_{n\to\infty} \int f_n,$$

which combined with Eq. (7.3) proves the theorem. ■

Remark 7.5 ("Explicit" Integral Formula). Given $f : \Omega \to [0, \infty]$ measurable, we know from the approximation Theorem 6.39 $\varphi_n \uparrow f$ where

$$\varphi_n := \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \leq \frac{k+1}{2^n}\right\}} + 2^n 1_{\{f > 2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned} \int_\Omega f d\mu &= \lim_{n\to\infty} \int_\Omega \varphi_n d\mu \\ &= \lim_{n\to\infty} \left[\sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} \mu\left(\frac{k}{2^n} < f \leq \frac{k+1}{2^n}\right) + 2^n \mu(f > 2^n) \right]. \end{aligned}$$

Corollary 7.6. *If $f_n \in L^+$ is a sequence of functions then*

$$\int \sum_{n=1}^\infty f_n = \sum_{n=1}^\infty \int f_n.$$

In particular, if $\sum_{n=1}^\infty \int f_n < \infty$ then $\sum_{n=1}^\infty f_n < \infty$ a.e.

Proof. First off we show that

$$\int (f_1 + f_2) = \int f_1 + \int f_2$$

by choosing non-negative simple function φ_n and ψ_n such that $\varphi_n \uparrow f_1$ and $\psi_n \uparrow f_2$. Then $(\varphi_n + \psi_n)$ is simple as well and $(\varphi_n + \psi_n) \uparrow (f_1 + f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int (f_1 + f_2) &= \lim_{n\to\infty} \int (\varphi_n + \psi_n) = \lim_{n\to\infty} \left(\int \varphi_n + \int \psi_n \right) \\ &= \lim_{n\to\infty} \int \varphi_n + \lim_{n\to\infty} \int \psi_n = \int f_1 + \int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum\limits_{n=1}^{N} f_n$ and $g = \sum\limits_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\begin{aligned}\sum_{n=1}^{\infty} \int f_n &:= \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n \\ &= \lim_{N\to\infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.\end{aligned}$$

■

Remark 7.7. It is in the proof of Corollary 7.6 (i.e. the linearity of the integral) that we really make use of the assumption that all of our functions are measurable. In fact the definition $\int f d\mu$ makes sense for **all** functions $f : \Omega \to [0, \infty]$ not just measurable functions. Moreover the monotone convergence theorem holds in this generality with no change in the proof. However, in the proof of Corollary 7.6, we use the approximation Theorem 6.39 which relies heavily on the measurability of the functions to be approximated.

Example 7.8 (Sums as Integrals I). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, $\mu(A) = \#(A)$ for $A \subset \Omega$ is the counting measure on $\mathcal{B}$, and $f : \mathbb{N} \to [0, \infty]$ is a function. Since

$$f = \sum_{n=1}^{\infty} f(n) 1_{\{n\}},$$

it follows from Corollary 7.6 that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^{\infty} \int_{\mathbb{N}} f(n) 1_{\{n\}} d\mu = \sum_{n=1}^{\infty} f(n) \mu(\{n\}) = \sum_{n=1}^{\infty} f(n).$$

Thus the integral relative to counting measure is simply the infinite sum.

Lemma 7.9 (Sums as Integrals II*). *Let Ω be a set and $\rho : \Omega \to [0, \infty]$ be a function, let $\mu = \sum_{\omega\in\Omega} \rho(\omega)\delta_\omega$ on $\mathcal{B} = 2^\Omega$, i.e.*

$$\mu(A) = \sum_{\omega\in A} \rho(\omega).$$

If $f : \Omega \to [0, \infty]$ is a function (which is necessarily measurable), then

$$\int_\Omega f d\mu = \sum_\Omega f\rho.$$

Proof. Suppose that $\varphi : \Omega \to [0, \infty)$ is a simple function, then $\varphi = \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}$ and

$$\begin{aligned}\sum_\Omega \varphi\rho &= \sum_{\omega\in\Omega} \rho(\omega) \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}(\omega) = \sum_{z\in[0,\infty)} z \sum_{\omega\in\Omega} \rho(\omega) 1_{\{\varphi=z\}}(\omega) \\ &= \sum_{z\in[0,\infty)} z\mu(\{\varphi = z\}) = \int_\Omega \varphi d\mu.\end{aligned}$$

So if $\varphi : \Omega \to [0, \infty)$ is a simple function such that $\varphi \le f$, then

$$\int_\Omega \varphi d\mu = \sum_\Omega \varphi\rho \le \sum_\Omega f\rho.$$

Taking the sup over φ in this last equation then shows that

$$\int_\Omega f d\mu \le \sum_\Omega f\rho.$$

For the reverse inequality, let $\Lambda \subset\subset \Omega$ be a finite set and $N \in (0, \infty)$. Set $f^N(\omega) = \min\{N, f(\omega)\}$ and let $\varphi_{N,\Lambda}$ be the simple function given by $\varphi_{N,\Lambda}(\omega) := 1_\Lambda(\omega) f^N(\omega)$. Because $\varphi_{N,\Lambda}(\omega) \le f(\omega)$,

$$\sum_\Lambda f^N \rho = \sum_\Omega \varphi_{N,\Lambda}\rho = \int_\Omega \varphi_{N,\Lambda} d\mu \le \int_\Omega f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_\Lambda f\rho \le \int_\Omega f d\mu.$$

Since Λ is arbitrary, this implies

$$\sum_\Omega f\rho \le \int_\Omega f d\mu.$$

■

Exercise 7.1. Suppose that $\mu_n : \mathcal{B} \to [0,\infty]$ are measures on $\mathcal{B}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{B}$. Prove that $\mu : \mathcal{B} \to [0,\infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Proposition 7.10. *Suppose that $f \geq 0$ is a measurable function. Then $\int_\Omega f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \geq 0$ are measurable functions such that $f \leq g$ a.e. then $\int f d\mu \leq \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

Proof. If $f = 0$ a.e. and $\varphi \leq f$ is a simple function then $\varphi = 0$ a.e. This implies that $\mu(\varphi^{-1}(\{y\})) = 0$ for all $y > 0$ and hence $\int_\Omega \varphi d\mu = 0$ and therefore $\int_\Omega f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 7.3),

$$\mu(f \geq 1/n) \leq n \int f d\mu = 0 \text{ for all } n.$$

Therefore, $\mu(f > 0) \leq \sum_{n=1}^\infty \mu(f \geq 1/n) = 0$, i.e. $f = 0$ a.e. For the second assertion let E be the exceptional set where $f > g$, i.e. $E := \{\omega \in \Omega : f(\omega) > g(\omega)\}$. By assumption E is a null set and $1_{E^c} f \leq 1_{E^c} g$ everywhere. Because $g = 1_{E^c} g + 1_E g$ and $1_E g = 0$ a.e.,

$$\int g d\mu = \int 1_{E^c} g d\mu + \int 1_E g d\mu = \int 1_{E^c} g d\mu$$

and similarly $\int f d\mu = \int 1_{E^c} f d\mu$. Since $1_{E^c} f \leq 1_{E^c} g$ everywhere,

$$\int f d\mu = \int 1_{E^c} f d\mu \leq \int 1_{E^c} g d\mu = \int g d\mu.$$

■

Corollary 7.11. *Suppose that $\{f_n\}$ is a sequence of non-negative measurable functions and f is a measurable function such that $f_n \uparrow f$ off a null set, then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Let $E \subset \Omega$ be a null set such that $f_n 1_{E^c} \uparrow f 1_{E^c}$ as $n \to \infty$. Then by the monotone convergence theorem and Proposition 7.10,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

Lemma 7.12 (Fatou's Lemma). *If $f_n : \Omega \to [0,\infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n$$

Proof. Define $g_k := \inf_{n \geq k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \leq f_n$ for all $k \leq n$,

$$\int g_k \leq \int f_n \text{ for all } n \geq k$$

and therefore

$$\int g_k \leq \lim \inf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \lim \inf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \leq \lim \inf_{n\to\infty} \int f_n.$$

■

The following Lemma and the next Corollary are simple applications of Corollary 7.6.

Lemma 7.13 (The First Borell – Cantelli Lemma). *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

$$\{A_n \text{ i.o.}\} = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^\infty \bigcup_{n \geq N} A_n.$$

If $\sum_{n=1}^\infty \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.

Proof. (First Proof.) Let us first observe that

$$\{A_n \text{ i.o.}\} = \left\{\omega \in \Omega : \sum_{n=1}^\infty 1_{A_n}(\omega) = \infty\right\}.$$

Hence if $\sum_{n=1}^\infty \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^\infty \mu(A_n) = \sum_{n=1}^\infty \int_\Omega 1_{A_n} d\mu = \int_\Omega \sum_{n=1}^\infty 1_{A_n} d\mu$$

implies that $\sum_{n=1}^\infty 1_{A_n}(\omega) < \infty$ for μ - a.e. ω. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$.
(Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned}\mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n \geq N} A_n\right) \\ &\leq \lim_{N\to\infty} \sum_{n \geq N} \mu(A_n)\end{aligned}$$

and the last limit is zero since $\sum_{n=1}^\infty \mu(A_n) < \infty$. ■

Corollary 7.14. *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{B}$ is a collection of sets such that $\mu(A_i \cap A_j) = 0$ for all $i \neq j$, then*

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n).$$

Proof. Since

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \int_\Omega 1_{\cup_{n=1}^\infty A_n} d\mu \text{ and}$$
$$\sum_{n=1}^\infty \mu(A_n) = \int_\Omega \sum_{n=1}^\infty 1_{A_n} d\mu$$

it suffices to show

$$\sum_{n=1}^\infty 1_{A_n} = 1_{\cup_{n=1}^\infty A_n} \ \mu - \text{a.e.} \tag{7.5}$$

Now $\sum_{n=1}^\infty 1_{A_n} \geq 1_{\cup_{n=1}^\infty A_n}$ and $\sum_{n=1}^\infty 1_{A_n}(\omega) \neq 1_{\cup_{n=1}^\infty A_n}(\omega)$ iff $\omega \in A_i \cap A_j$ for some $i \neq j$, that is

$$\left\{\omega : \sum_{n=1}^\infty 1_{A_n}(\omega) \neq 1_{\cup_{n=1}^\infty A_n}(\omega)\right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (7.5) and hence the corollary. ∎

7.2 Integrals of Complex Valued Functions

Definition 7.15. *A measurable function $f : \Omega \to \bar{\mathbb{R}}$ is **integrable** if $f_+ := f 1_{\{f \geq 0\}}$ and $f_- = -f \ 1_{\{f \leq 0\}}$ are **integrable**. We write $\mathrm{L}^1(\mu; \mathbb{R})$ for the space of real valued integrable functions. For $f \in \mathrm{L}^1(\mu; \mathbb{R})$, let*

$$\int_\Omega f d\mu = \int_\Omega f_+ d\mu - \int_\Omega f_- d\mu.$$

To shorten notation in this chapter we may simply write $\int f d\mu$ or even $\int f$ for $\int_\Omega f d\mu$.

Convention: If $f, g : \Omega \to \bar{\mathbb{R}}$ are two measurable functions, let $f + g$ denote the collection of measurable functions $h : \Omega \to \bar{\mathbb{R}}$ such that $h(\omega) = f(\omega) + g(\omega)$ whenever $f(\omega) + g(\omega)$ is well defined, i.e. is not of the form $\infty - \infty$ or $-\infty + \infty$. We use a similar convention for $f - g$. Notice that if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $h_1, h_2 \in f + g$, then $h_1 = h_2$ a.e. because $|f| < \infty$ and $|g| < \infty$ a.e.

Notation 7.16 (Abuse of notation) *We will sometimes denote the integral $\int_\Omega f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{B}$.*

Remark 7.17. Since

$$f_\pm \leq |f| \leq f_+ + f_-,$$

a measurable function f is **integrable** iff $\int |f| \ d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu; \mathbb{R}) := \left\{f : \Omega \to \bar{\mathbb{R}} : \ f \text{ is measurable and } \int_\Omega |f| \ d\mu < \infty\right\}.$$

If $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $f = g$ a.e. then $f_\pm = g_\pm$ a.e. and so it follows from Proposition 7.10 that $\int f d\mu = \int g d\mu$. In particular if $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ we may define

$$\int_\Omega (f + g)\, d\mu = \int_\Omega h d\mu$$

where h is any element of $f + g$.

Proposition 7.18. *The map*

$$f \in \mathrm{L}^1(\mu; \mathbb{R}) \to \int_\Omega f d\mu \in \mathbb{R}$$

is linear and has the monotonicity property: $\int f d\mu \leq \int g d\mu$ for all $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ such that $f \leq g$ a.e.

Proof. Let $f, g \in \mathrm{L}^1(\mu; \mathbb{R})$ and $a, b \in \mathbb{R}$. By modifying f and g on a null set, we may assume that f, g are real valued functions. We have $af + bg \in \mathrm{L}^1(\mu; \mathbb{R})$ because

$$|af + bg| \leq |a| \, |f| + |b| \, |g| \in \mathrm{L}^1(\mu; \mathbb{R}).$$

If $a < 0$, then

$$(af)_+ = -af_- \text{ and } (af)_- = -af_+$$

so that

$$\int af = -a \int f_- + a \int f_+ = a\left(\int f_+ - \int f_-\right) = a \int f.$$

A similar calculation works for $a > 0$ and the case $a = 0$ is trivial so we have shown that

$$\int af = a \int f.$$

Now set $h = f + g$. Since $h = h_+ - h_-$,

$$h_+ - h_- = f_+ - f_- + g_+ - g_-$$

or

$$h_+ + f_- + g_- = h_- + f_+ + g_+.$$

Therefore,

$$\int h_+ + \int f_- + \int g_- = \int h_- + \int f_+ + \int g_+$$

and hence

$$\int h = \int h_+ - \int h_- = \int f_+ + \int g_+ - \int f_- - \int g_- = \int f + \int g.$$

Finally if $f_+ - f_- = f \le g = g_+ - g_-$ then $f_+ + g_- \le g_+ + f_-$ which implies that

$$\int f_+ + \int g_- \le \int g_+ + \int f_-$$

or equivalently that

$$\int f = \int f_+ - \int f_- \le \int g_+ - \int g_- = \int g.$$

The monotonicity property is also a consequence of the linearity of the integral, the fact that $f \le g$ a.e. implies $0 \le g - f$ a.e. and Proposition 7.10. ■

Definition 7.19. *A measurable function $f : \Omega \to \mathbb{C}$ is* ***integrable*** *if $\int_\Omega |f| \; d\mu < \infty$. Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : \Omega \to \mathbb{C} : \ f \text{ is measurable and } \int_\Omega |f| \; d\mu < \infty \right\}.$$

denote the complex valued integrable functions. Because, $\max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|) \le |f| \le \sqrt{2} \max(|\mathrm{Re}\, f|, |\mathrm{Im}\, f|)$, *$\int |f| \; d\mu < \infty$ iff*

$$\int |\mathrm{Re}\, f| \, d\mu + \int |\mathrm{Im}\, f| \, d\mu < \infty.$$

For $f \in \mathrm{L}^1(\mu;\mathbb{C})$ define

$$\int f \; d\mu = \int \mathrm{Re}\, f \; d\mu + i \int \mathrm{Im}\, f \; d\mu.$$

It is routine to show the integral is still linear on $\mathrm{L}^1(\mu;\mathbb{C})$ (prove!). In the remainder of this section, let $\mathrm{L}^1(\mu)$ be either $\mathrm{L}^1(\mu;\mathbb{C})$ or $\mathrm{L}^1(\mu;\mathbb{R})$. If $A \in \mathcal{B}$ and $f \in \mathrm{L}^1(\mu;\mathbb{C})$ or $f : \Omega \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_\Omega 1_A f d\mu.$$

Proposition 7.20. *Suppose that $f \in \mathrm{L}^1(\mu;\mathbb{C})$, then*

$$\left| \int_\Omega f d\mu \right| \le \int_\Omega |f| \, d\mu. \tag{7.6}$$

Proof. Start by writing $\int_\Omega f \; d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_\Omega f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta} \int_\Omega f \; d\mu = \int_\Omega e^{-i\theta} f \; d\mu = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i \int_\Omega \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_\Omega \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 7.10,

$$\left| \int_\Omega f d\mu \right| = \int_\Omega \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_\Omega \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_\Omega |f| \, d\mu.$$

■

Proposition 7.21. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \ne 0\}$ is σ – finite, in fact $\{|f| \ge \frac{1}{n}\} \uparrow \{f \ne 0\}$ and $\mu(|f| \ge \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) *$\int_E f = \int_E g$ for all $E \in \mathcal{B}$*
   b) *$\int\limits_\Omega |f - g| = 0$*
   c) *$f = g$ a.e.*

Proof. 1. By Chebyshev's inequality, Lemma 7.3,

$$\mu(|f| \ge \frac{1}{n}) \le n \int_\Omega |f| \, d\mu < \infty$$

for all n.

2. (a) $\Longrightarrow$ (c) Notice that

$$\int_E f = \int_E g \Leftrightarrow \int_E (f - g) = 0$$

for all $E \in \mathcal{B}$. Taking $E = \{\mathrm{Re}(f - g) > 0\}$ and using $1_E \mathrm{Re}(f - g) \ge 0$, we learn that

$$0 = \mathrm{Re} \int_E (f - g) d\mu = \int 1_E \mathrm{Re}(f - g) \Longrightarrow 1_E \mathrm{Re}(f - g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

$$\mu(\{\text{Re}(f-g)>0\}) = \mu(E) = 0.$$

Similar $\mu(\text{Re}(f-g)<0) = 0$ so that $\text{Re}(f-g) = 0$ a.e. Similarly, $\text{Im}(f-g) = 0$ a.e and hence $f - g = 0$ a.e., i.e. $f = g$ a.e.

(c) $\Longrightarrow$ (b) is clear and so is (b) $\Longrightarrow$ (a) since

$$\left|\int_E f - \int_E g\right| \leq \int |f-g| = 0.$$

■

Lemma 7.22. *Suppose that $h \in \mathrm{L}^1(\mu)$ satisfies*

$$\int_A h d\mu \geq 0 \text{ for all } A \in \mathcal{B}, \tag{7.7}$$

then $h \geq 0$ a.e.

Proof. Since by assumption,

$$0 = \text{Im} \int_A h d\mu = \int_A \text{Im}\, h d\mu \text{ for all } A \in \mathcal{B},$$

we may apply Proposition 7.21 to conclude that $\text{Im}\, h = 0$ a.e. Thus we may now assume that h is real valued. Taking $A = \{h < 0\}$ in Eq. (7.7) implies

$$\int_\Omega 1_A |h| \, d\mu = \int_\Omega -1_A h d\mu = -\int_A h d\mu \leq 0.$$

However $1_A |h| \geq 0$ and therefore it follows that $\int_\Omega 1_A |h| \, d\mu = 0$ and so Proposition 7.21 implies $1_A |h| = 0$ a.e. which then implies $0 = \mu(A) = \mu(h<0) = 0$.
■

Lemma 7.23 (Integral Comparison). *Suppose $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space (i.e. there exists $\Omega_n \in \mathcal{B}$ such that $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n) < \infty$ for all n) and $f, g : \Omega \to [0, \infty]$ are $\mathcal{B}$ – measurable functions. Then $f \geq g$ a.e. iff*

$$\int_A f d\mu \geq \int_A g d\mu \text{ for all } A \in \mathcal{B}. \tag{7.8}$$

In particular $f = g$ a.e. iff equality holds in Eq. (7.8).

Proof. It was already shown in Proposition 7.10 that $f \geq g$ a.e. implies Eq. (7.8). For the converse assertion, let $B_n := \{f \leq n 1_{\Omega_n}\}$. Then from Eq. (7.8),

$$\infty > n\mu(\Omega_n) \geq \int f 1_{B_n} d\mu \geq \int g 1_{B_n} d\mu$$

from which it follows that both $f 1_{B_n}$ and $g 1_{B_n}$ are in $\mathrm{L}^1(\mu)$ and hence $h := f 1_{B_n} - g 1_{B_n} \in \mathrm{L}^1(\mu)$. Using Eq. (7.8) again we know that

$$\int_A h = \int f 1_{B_n \cap A} - \int g 1_{B_n \cap A} \geq 0 \text{ for all } A \in \mathcal{B}.$$

An application of Lemma 7.22 implies $h \geq 0$ a.e., i.e. $f 1_{B_n} \geq g 1_{B_n}$ a.e. Since $B_n \uparrow \{f < \infty\}$, we may conclude that

$$f 1_{\{f<\infty\}} = \lim_{n\to\infty} f 1_{B_n} \geq \lim_{n\to\infty} g 1_{B_n} = g 1_{\{f<\infty\}} \text{ a.e.}$$

Since $f \geq g$ whenever $f = \infty$, we have shown $f \geq g$ a.e.

If equality holds in Eq. (7.8), then we know that $g \leq f$ and $f \leq g$ a.e., i.e. $f = g$ a.e. ■

Notice that we can not drop the σ – finiteness assumption in Lemma 7.23. For example, let μ be the measure on $\mathcal{B}$ such that $\mu(A) = \infty$ when $A \neq \emptyset$, $g = 3$, and $f = 2$. Then equality holds (both sides are infinite unless $A = \emptyset$ when they are both zero) in Eq. (7.8) holds even though $f < g$ **everywhere.**

Definition 7.24. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $L^1(\mu) = L^1(\Omega, \mathcal{B}, \mu)$ denote the set of $\mathrm{L}^1(\mu)$ functions modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e. We make this into a normed space using the norm*

$$\|f - g\|_{L^1} = \int |f - g| \, d\mu$$

and into a metric space using $\rho_1(f, g) = \|f - g\|_{L^1}$.

Warning: in the future we will often not make much of a distinction between $L^1(\mu)$ and $\mathrm{L}^1(\mu)$. On occasion this can be dangerous and this danger will be pointed out when necessary.

Remark 7.25. More generally we may define $L^p(\mu) = L^p(\Omega, \mathcal{B}, \mu)$ for $p \in [1, \infty)$ as the set of measurable functions f such that

$$\int_\Omega |f|^p \, d\mu < \infty$$

modulo the equivalence relation; $f \sim g$ iff $f = g$ a.e.

We will see in later that

$$\|f\|_{L^p} = \left(\int |f|^p \, d\mu\right)^{1/p} \text{ for } f \in L^p(\mu)$$

is a norm and $(L^p(\mu), \|\cdot\|_{L^p})$ is a Banach space in this norm and in particular,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \text{ for all } f, g \in L^p(\mu).$$

Theorem 7.26 (Dominated Convergence Theorem). *Suppose* $f_n, g_n, g \in \mathrm{L}^1(\mu)$, $f_n \to f$ *a.e.*, $|f_n| \le g_n \in \mathrm{L}^1(\mu)$, $g_n \to g$ *a.e. and* $\int_\Omega g_n d\mu \to \int_\Omega g d\mu$. *Then* $f \in \mathrm{L}^1(\mu)$ *and*

$$\int_\Omega f d\mu = \lim_{h\to\infty} \int_\Omega f_n d\mu.$$

(In most typical applications of this theorem $g_n = g \in \mathrm{L}^1(\mu)$ *for all* n.*)*

Proof. Notice that $|f| = \lim_{n\to\infty} |f_n| \le \lim_{n\to\infty} |g_n| \le g$ a.e. so that $f \in \mathrm{L}^1(\mu)$. By considering the real and imaginary parts of f separately, it suffices to prove the theorem in the case where f is real. By Fatou's Lemma,

$$\begin{aligned}\int_\Omega (g \pm f) d\mu &= \int_\Omega \liminf_{n\to\infty} (g_n \pm f_n)\, d\mu \le \liminf_{n\to\infty} \int_\Omega (g_n \pm f_n)\, d\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu \right) \\ &= \int_\Omega g d\mu + \liminf_{n\to\infty} \left(\pm \int_\Omega f_n d\mu \right)\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_\Omega g d\mu \pm \int_\Omega f d\mu \le \int_\Omega g d\mu + \begin{cases} \liminf_{n\to\infty} \int_\Omega f_n d\mu \\ -\limsup_{n\to\infty} \int_\Omega f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_\Omega f_n d\mu \le \int_\Omega f d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

This shows that $\lim_{n\to\infty} \int_\Omega f_n d\mu$ exists and is equal to $\int_\Omega f d\mu$. ■

Exercise 7.2. Give another proof of Proposition 7.20 by first proving Eq. (7.6) with f being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 6.39 along with the dominated convergence Theorem 7.26 to handle the general case.

Corollary 7.27. *Let* $\{f_n\}_{n=1}^\infty \subset \mathrm{L}^1(\mu)$ *be a sequence such that* $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$, *then* $\sum_{n=1}^\infty f_n$ *is convergent a.e. and*

$$\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.$$

Proof. The condition $\sum_{n=1}^\infty \|f_n\|_{\mathrm{L}^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \le \sum_{n=1}^N |f_n| \le \sum_{n=1}^\infty |f_n| \in \mathrm{L}^1(\mu).$$

So by the dominated convergence theorem,

$$\begin{aligned}\int_\Omega \left(\sum_{n=1}^\infty f_n \right) d\mu &= \int_\Omega \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty} \int_\Omega S_N d\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int_\Omega f_n d\mu = \sum_{n=1}^\infty \int_\Omega f_n d\mu.\end{aligned}$$

■

Example 7.28 (Sums as integrals). Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, μ is counting measure on $\mathcal{B}$ (see Example 7.8), and $f : \mathbb{N} \to \mathbb{C}$ is a function. From Example 7.8 we have $f \in L^1(\mu)$ iff $\sum_{n=1}^\infty |f(n)| < \infty$, i.e. iff the sum, $\sum_{n=1}^\infty f(n)$ is absolutely convergent. Moreover, if $f \in L^1(\mu)$, we may again write

$$f = \sum_{n=1}^\infty f(n)\, 1_{\{n\}}$$

and then use Corollary 7.27 to conclude that

$$\int_{\mathbb{N}} f d\mu = \sum_{n=1}^\infty \int_{\mathbb{N}} f(n)\, 1_{\{n\}} d\mu = \sum_{n=1}^\infty f(n)\, \mu(\{n\}) = \sum_{n=1}^\infty f(n).$$

So again the integral relative to counting measure is simply the infinite sum **provided** the sum is absolutely convergent.

However if $f(n) = (-1)^n \frac{1}{n}$, then

$$\sum_{n=1}^\infty f(n) := \lim_{N\to\infty} \sum_{n=1}^N f(n)$$

is perfectly well defined while $\int_{\mathbb{N}} f d\mu$ is **not**. In fact in this case we have,

$$\int_{\mathbb{N}} f_\pm d\mu = \infty.$$

The point is that when we write $\sum_{n=1}^\infty f(n)$ the ordering of the terms in the sum may matter. On the other hand, $\int_{\mathbb{N}} f d\mu$ knows nothing about the integer ordering.

The following corollary will be routinely be used in the sequel – often without explicit mention.

Corollary 7.29 (Differentiation Under the Integral). *Suppose that* $J \subset \mathbb{R}$ *is an open interval and* $f : J \times \Omega \to \mathbb{C}$ *is a function such that*

1. $\omega \to f(t, \omega)$ *is measurable for each* $t \in J$.
2. $f(t_0, \cdot) \in L^1(\mu)$ *for some* $t_0 \in J$.
3. $\frac{\partial f}{\partial t}(t, \omega)$ *exists for all* (t, ω).
4. *There is a function* $g \in \mathrm{L}^1(\mu)$ *such that* $\left|\frac{\partial f}{\partial t}(t, \cdot)\right| \leq g$ *for each* $t \in J$.

Then $f(t, \cdot) \in \mathrm{L}^1(\mu)$ *for all* $t \in J$ *(i.e.* $\int_\Omega |f(t, \omega)| \, d\mu(\omega) < \infty$*),* $t \to \int_\Omega f(t, \omega) d\mu(\omega)$ *is a differentiable function on* J*, and*

$$\frac{d}{dt} \int_\Omega f(t, \omega) d\mu(\omega) = \int_\Omega \frac{\partial f}{\partial t}(t, \omega) d\mu(\omega).$$

Proof. By considering the real and imaginary parts of f separately, we may assume that f is real. Also notice that

$$\frac{\partial f}{\partial t}(t, \omega) = \lim_{n \to \infty} n(f(t + n^{-1}, \omega) - f(t, \omega))$$

and therefore, for $\omega \to \frac{\partial f}{\partial t}(t, \omega)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t, \omega) - f(t_0, \omega)| \leq g(\omega) \, |t - t_0| \text{ for all } t \in J \tag{7.9}$$

and hence

$$|f(t, \omega)| \leq |f(t, \omega) - f(t_0, \omega)| + |f(t_0, \omega)| \leq g(\omega) \, |t - t_0| + |f(t_0, \omega)| \, .$$

This shows $f(t, \cdot) \in \mathrm{L}^1(\mu)$ for all $t \in J$. Let $G(t) := \int_\Omega f(t, \omega) d\mu(\omega)$, then

$$\frac{G(t) - G(t_0)}{t - t_0} = \int_\Omega \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} d\mu(\omega).$$

By assumption,

$$\lim_{t \to t_0} \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} = \frac{\partial f}{\partial t}(t, \omega) \text{ for all } \omega \in \Omega$$

and by Eq. (7.9),

$$\left|\frac{f(t, \omega) - f(t_0, \omega)}{t - t_0}\right| \leq g(\omega) \text{ for all } t \in J \text{ and } \omega \in \Omega.$$

Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned} \lim_{n \to \infty} \frac{G(t_n) - G(t_0)}{t_n - t_0} &= \lim_{n \to \infty} \int_\Omega \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} d\mu(\omega) \\ &= \int_\Omega \lim_{n \to \infty} \frac{f(t_n, \omega) - f(t_0, \omega)}{t_n - t_0} d\mu(\omega) \\ &= \int_\Omega \frac{\partial f}{\partial t}(t_0, \omega) d\mu(\omega) \end{aligned}$$

for **all** sequences $t_n \in J \setminus \{t_0\}$ such that $t_n \to t_0$. Therefore, $\dot{G}(t_0) = \lim_{t \to t_0} \frac{G(t) - G(t_0)}{t - t_0}$ exists and

$$\dot{G}(t_0) = \int_\Omega \frac{\partial f}{\partial t}(t_0, \omega) d\mu(\omega).$$

■

Corollary 7.30. *Suppose that* $\{a_n\}_{n=0}^\infty \subset \mathbb{C}$ *is a sequence of complex numbers such that series*

$$f(z) := \sum_{n=0}^\infty a_n (z - z_0)^n$$

is convergent for $|z - z_0| < R$*, where* R *is some positive number. Then* $f : D(z_0, R) \to \mathbb{C}$ *is complex differentiable on* $D(z_0, R)$ *and*

$$f'(z) = \sum_{n=0}^\infty n a_n (z - z_0)^{n-1} = \sum_{n=1}^\infty n a_n (z - z_0)^{n-1}. \tag{7.10}$$

By induction it follows that $f^{(k)}$ *exists for all* k *and that*

$$f^{(k)}(z) = \sum_{n=0}^\infty n(n-1) \dots (n-k+1) a_n (z - z_0)^{n-1}.$$

Proof. Let $\rho < R$ be given and choose $r \in (\rho, R)$. Since $z = z_0 + r \in D(z_0, R)$, by assumption the series $\sum_{n=0}^\infty a_n r^n$ is convergent and in particular $M := \sup_n |a_n r^n| < \infty$. We now apply Corollary 7.29 with $X = \mathbb{N} \cup \{0\}$, μ being counting measure, $\Omega = D(z_0, \rho)$ and $g(z, n) := a_n (z - z_0)^n$. Since

$$\begin{aligned} |g'(z, n)| &= |n a_n (z - z_0)^{n-1}| \leq n \, |a_n| \, \rho^{n-1} \\ &\leq \frac{1}{r} n \left(\frac{\rho}{r}\right)^{n-1} |a_n| \, r^n \leq \frac{1}{r} n \left(\frac{\rho}{r}\right)^{n-1} M \end{aligned}$$

and the function $G(n) := \frac{M}{r} n \left(\frac{\rho}{r}\right)^{n-1}$ is summable (by the Ratio test for example), we may use G as our dominating function. It then follows from Corollary 7.29

$$f(z) = \int_X g(z,n) d\mu(n) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

is complex differentiable with the differential given as in Eq. (7.10). ∎

Definition 7.31 (Moment Generating Function). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. The* ***moment generating function*** *of X is $M_X : \mathbb{R} \to [0,\infty]$ defined by*

$$M_X(t) := \mathbb{E}\left[e^{tX}\right].$$

Proposition 7.32. *Suppose there exists $\varepsilon > 0$ such that $\mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty$, then $M_X(t)$ is a smooth function of $t \in (-\varepsilon, \varepsilon)$ and*

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n \text{ if } |t| \le \varepsilon. \tag{7.11}$$

In particular,

$$\mathbb{E}X^n = \left(\frac{d}{dt}\right)^n |_{t=0} M_X(t) \text{ for all } n \in \mathbb{N}_0. \tag{7.12}$$

Proof. If $|t| \le \varepsilon$, then

$$\mathbb{E}\left[\sum_{n=0}^{\infty} \frac{|t|^n}{n!} |X|^n\right] \le \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{\varepsilon^n}{n!} |X|^n\right] = \mathbb{E}\left[e^{\varepsilon|X|}\right] < \infty.$$

it $e^{tX} \le e^{\varepsilon|X|}$ for all $|t| \le \varepsilon$. Hence it follows from Corollary 7.29 that, for $|t| \le \varepsilon$,

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{t^n}{n!} X^n\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n.$$

Equation (7.12) now is a consequence of Corollary 7.30. ∎

7.2.1 Square Integrable Random Variables and Correlations

Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space. We say that $X : \Omega \to \mathbb{R}$ is **integrable** if $X \in L^1(P)$ and **square integrable** if $X \in L^2(P)$. When X is integrable we let $a_X := \mathbb{E}X$ be the **mean** of X.

Now suppose that $X, Y : \Omega \to \mathbb{R}$ are two square integrable random variables. Since

$$0 \le |X - Y|^2 = |X|^2 + |Y|^2 - 2|X||Y|,$$

it follows that

$$|XY| \le \frac{1}{2}|X|^2 + \frac{1}{2}|Y|^2 \in L^1(P).$$

In particular by taking $Y = 1$, we learn that $|X| \le \frac{1}{2}\left(1 + \left|X^2\right|\right)$ which shows that every square integrable random variable is also integrable.

Definition 7.33. *The* ***covariance,*** $\mathrm{Cov}(X,Y)$, *of two square integrable random variables, X and Y, is defined by*

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - a_X)(Y - a_Y)\right] = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

where $a_X := \mathbb{E}X$ and $a_Y := \mathbb{E}Y$. The ***variance*** *of X,*

$$\mathrm{Var}(X) := \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \tag{7.13}$$

We say that X and Y are ***uncorrelated*** *if $\mathrm{Cov}(X,Y) = 0$, i.e. $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$. More generally we say $\{X_k\}_{k=1}^n \subset L^2(P)$ are* ***uncorrelated*** *iff $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \ne j$.*

It follows from Eq. (7.13) that

$$\mathrm{Var}(X) \le \mathbb{E}\left[X^2\right] \text{ for all } X \in L^2(P). \tag{7.14}$$

Lemma 7.34. *The covariance function,* $\mathrm{Cov}(X,Y)$ *is bilinear in X and Y and $\mathrm{Cov}(X,Y) = 0$ if either X or Y is constant. For any constant k, $\mathrm{Var}(X+k) = \mathrm{Var}(X)$ and $\mathrm{Var}(kX) = k^2\,\mathrm{Var}(X)$. If $\{X_k\}_{k=1}^n$ are uncorrelated $L^2(P)$ – random variables, then*

$$\mathrm{Var}(S_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

Proof. We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove $\mathrm{Var}(X+k) = \mathrm{Var}(X)$;

$$\begin{aligned}\mathrm{Var}(X+k) &= \mathrm{Cov}(X+k, X+k) = \mathrm{Cov}(X+k, X) + \mathrm{Cov}(X+k, k) \\ &= \mathrm{Cov}(X+k, X) = \mathrm{Cov}(X,X) + \mathrm{Cov}(k,X) \\ &= \mathrm{Cov}(X,X) = \mathrm{Var}(X),\end{aligned}$$

wherein we have used the bilinearity of $\mathrm{Cov}(\cdot,\cdot)$ and the property that $\mathrm{Cov}(Y,k) = 0$ whenever k is a constant. ∎

Exercise 7.3 (A Weak Law of Large Numbers). Assume $\{X_n\}_{n=1}^{\infty}$ is a sequence if uncorrelated square integrable random variables which are identically distributed, i.e. $X_n \overset{d}{=} X_m$ for all $m, n \in \mathbb{N}$. Let $S_n := \sum_{k=1}^n X_k$, $\mu := \mathbb{E}X_k$ and $\sigma^2 := \mathrm{Var}(X_k)$ (these are independent of k). Show;

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu,$$
$$\mathbb{E}\left(\frac{S_n}{n} - \mu\right)^2 = \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}, \text{ and}$$
$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$

for all $\varepsilon > 0$ and $n \in \mathbb{N}$. (Compare this with Exercise 4.13.)

7.2.2 Some Discrete Distributions

Definition 7.35 (Generating Function). *Suppose that $N : \Omega \to \mathbb{N}_0$ is an integer valued random variable on a probability space, $(\Omega, \mathcal{B}, P)$. The generating function associated to N is defined by*

$$G_N(z) := \mathbb{E}\left[z^N\right] = \sum_{n=0}^{\infty} P(N = n) z^n \text{ for } |z| \le 1. \tag{7.15}$$

By Corollary 7.30, it follows that $P(N = n) = \frac{1}{n!} G_N^{(n)}(0)$ so that G_N can be used to completely recover the distribution of N.

Proposition 7.36 (Generating Functions). *The generating function satisfies,*

$$G_N^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right] \text{ for } |z| < 1$$

and

$$G^{(k)}(1) = \lim_{z \uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right],$$

where it is possible that one and hence both sides of this equation are infinite. In particular, $G'(1) := \lim_{z\uparrow 1} G'(z) = \mathbb{E}N$ and if $\mathbb{E}N^2 < \infty$,

$$\mathrm{Var}(N) = G''(1) + G'(1) - \left[G'(1)\right]^2. \tag{7.16}$$

Proof. By Corollary 7.30 for $|z| < 1$,

$$\begin{aligned} G_N^{(k)}(z) &= \sum_{n=0}^{\infty} P(N = n) \cdot n(n-1)\dots(n-k+1) z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1) z^{N-k}\right]. \end{aligned} \tag{7.17}$$

Since, for $z \in (0, 1)$,

$$0 \le N(N-1)\dots(N-k+1) z^{N-k} \uparrow N(N-1)\dots(N-k+1) \text{ as } z \uparrow 1,$$

we may apply the MCT to pass to the limit as $z \uparrow 1$ in Eq. (7.17) to find,

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right].$$

■

Exercise 7.4 (Some Discrete Distributions). Let $p \in (0, 1]$ and $\lambda > 0$. In the four parts below, the distribution of N will be described. You should work out the generating function, $G_N(z)$, in each case and use it to verify the given formulas for $\mathbb{E}N$ and $\mathrm{Var}(N)$.

1. Bernoulli(p) : $P(N = 1) = p$ and $P(N = 0) = 1 - p$. You should find $\mathbb{E}N = p$ and $\mathrm{Var}(N) = p - p^2$.
2. Binomial(n, p) : $P(N = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \dots, n$. ($P(N = k)$ is the probability of k successes in a sequence of n independent yes/no experiments with probability of success being p.) You should find $\mathbb{E}N = np$ and $\mathrm{Var}(N) = n\left(p - p^2\right)$.
3. Geometric(p) : $P(N = k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}$. ($P(N = k)$ is the probability that the k^{th} – trial is the first time of success out a sequence of independent trials with probability of success being p.) You should find $\mathbb{E}N = 1/p$ and $\mathrm{Var}(N) = \frac{1-p}{p^2}$.
4. Poisson(λ) : $P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ for all $k \in \mathbb{N}_0$. (We will come back to the interpretation of this distribution later.) You should find $\mathbb{E}N = \lambda = \mathrm{Var}(N)$.

7.3 Integration on $\mathbb{R}$

Notation 7.37 *If m is Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$, f is a non-negative Borel measurable function and $a < b$ with $a, b \in \bar{\mathbb{R}}$, we will often write $\int_a^b f(x)\,dx$ or $\int_a^b f dm$ for $\int_{(a,b]\cap\mathbb{R}} f dm$.*

Example 7.38. Suppose $-\infty < a < b < \infty$, $f \in C([a, b], \mathbb{R})$ and m be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi = \{a = a_0 < a_1 < \dots < a_n = b\},$$

let

$$\mathrm{mesh}(\pi) := \max\{|a_j - a_{j-1}| : j = 1, \dots, n\}$$

and

$$f_\pi(x) := \sum_{l=0}^{n-1} f(a_l) 1_{(a_l, a_{l+1}]}(x).$$

Then

$$\int_a^b f_\pi \, dm = \sum_{l=0}^{n-1} f(a_l)\, m((a_l, a_{l+1}]) = \sum_{l=0}^{n-1} f(a_l)(a_{l+1} - a_l)$$

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty} \mathrm{mesh}(\pi_k) = 0$, we know that

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b f(x)\, dx \tag{7.18}$$

where the latter integral is the Riemann integral. Using the (uniform) continuity of f on $[a,b]$, it easily follows that $\lim_{k\to\infty} f_{\pi_k}(x) = f(x)$ and that $|f_{\pi_k}(x)| \le g(x) := M 1_{(a,b]}(x)$ for all $x \in (a,b]$ where $M := \max_{x\in[a,b]} |f(x)| < \infty$. Since $\int_{\mathbb{R}} g dm = M(b-a) < \infty$, we may apply D.C.T. to conclude,

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b \lim_{k\to\infty} f_{\pi_k} \, dm = \int_a^b f \, dm.$$

This equation with Eq. (7.18) shows

$$\int_a^b f \, dm = \int_a^b f(x)\, dx$$

whenever $f \in C([a,b],\mathbb{R})$, i.e. the Lebesgue and the Riemann integral agree on continuous functions. See Theorem 7.61 below for a more general statement along these lines.

Theorem 7.39 (The Fundamental Theorem of Calculus). *Suppose* $-\infty < a < b < \infty$, $f \in C((a,b),\mathbb{R}) \cap L^1((a,b),m)$ *and* $F(x) := \int_a^x f(y) dm(y)$. *Then*

1. $F \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$.
2. $F'(x) = f(x)$ *for all* $x \in (a,b)$.
3. *If* $G \in C([a,b],\mathbb{R}) \cap C^1((a,b),\mathbb{R})$ *is an anti-derivative of* f *on* (a,b) *(i.e.* $f = G'|_{(a,b)}$*) then*
$$\int_a^b f(x) dm(x) = G(b) - G(a).$$

Proof. Since $F(x) := \int_{\mathbb{R}} 1_{(a,x)}(y) f(y) dm(y)$, $\lim_{x\to z} 1_{(a,x)}(y) = 1_{(a,z)}(y)$ for m – a.e. y and $\left|1_{(a,x)}(y) f(y)\right| \le 1_{(a,b)}(y) |f(y)|$ is an L^1 – function, it follows from the dominated convergence Theorem 7.26 that F is continuous on $[a,b]$. Simple manipulations show,

$$\left|\frac{F(x+h) - F(x)}{h} - f(x)\right| = \frac{1}{|h|} \begin{cases} \left|\int_x^{x+h} [f(y) - f(x)]\, dm(y)\right| & \text{if } h > 0 \\ \left|\int_{x+h}^x [f(y) - f(x)]\, dm(y)\right| & \text{if } h < 0 \end{cases}$$
$$\le \frac{1}{|h|} \begin{cases} \int_x^{x+h} |f(y) - f(x)|\, dm(y) & \text{if } h > 0 \\ \int_{x+h}^x |f(y) - f(x)|\, dm(y) & \text{if } h < 0 \end{cases}$$
$$\le \sup\{|f(y) - f(x)| : y \in [x - |h|, x + |h|]\}$$

and the latter expression, by the continuity of f, goes to zero as $h \to 0$. This shows $F' = f$ on (a,b).

For the converse direction, we have by assumption that $G'(x) = F'(x)$ for $x \in (a,b)$. Therefore by the mean value theorem, $F - G = C$ for some constant C. Hence

$$\int_a^b f(x) dm(x) = F(b) = F(b) - F(a)$$
$$= (G(b) + C) - (G(a) + C) = G(b) - G(a).$$

■

We can use the above results to integrate some non-Riemann integrable functions:

Example 7.40. For all $\lambda > 0$,

$$\int_0^\infty e^{-\lambda x} dm(x) = \lambda^{-1} \text{ and } \int_{\mathbb{R}} \frac{1}{1+x^2} dm(x) = \pi.$$

The proof of these identities are similar. By the monotone convergence theorem, Example 7.38 and the fundamental theorem of calculus for Riemann integrals (or Theorem 7.39 below),

$$\int_0^\infty e^{-\lambda x} dm(x) = \lim_{N\to\infty} \int_0^N e^{-\lambda x} dm(x) = \lim_{N\to\infty} \int_0^N e^{-\lambda x} dx$$
$$= -\lim_{N\to\infty} \frac{1}{\lambda} e^{-\lambda x}|_0^N = \lambda^{-1}$$

and

$$\int_{\mathbb{R}} \frac{1}{1+x^2} dm(x) = \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dm(x) = \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dx$$
$$= \lim_{N\to\infty} \left[\tan^{-1}(N) - \tan^{-1}(-N)\right] = \pi.$$

Let us also consider the functions x^{-p},

$$\int_{(0,1]} \frac{1}{x^p}\, dm(x) = \lim_{n\to\infty} \int_0^1 1_{(\frac{1}{n},1]}(x) \frac{1}{x^p} dm(x)$$
$$= \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x^p} dx = \lim_{n\to\infty} \frac{x^{-p+1}}{1-p}\Big|_{1/n}^1$$
$$= \begin{cases} \frac{1}{1-p} & \text{if } p<1 \\ \infty & \text{if } p>1 \end{cases}$$

If $p = 1$ we find

$$\int_{(0,1]} \frac{1}{x^p}\, dm(x) = \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x} dx = \lim_{n\to\infty} \ln(x)|_{1/n}^1 = \infty.$$

Exercise 7.5. Show

$$\int_1^\infty \frac{1}{x^p} dm\,(x) = \begin{cases} \infty & \text{if } p \le 1 \\ \frac{1}{p-1} & \text{if } p > 1 \end{cases}.$$

Example 7.41 (Integration of Power Series). Suppose $R > 0$ and $\{a_n\}_{n=0}^\infty$ is a sequence of complex numbers such that $\sum_{n=0}^\infty |a_n|\, r^n < \infty$ for all $r \in (0, R)$. Then

$$\int_\alpha^\beta \left(\sum_{n=0}^\infty a_n x^n\right) dm(x) = \sum_{n=0}^\infty a_n \int_\alpha^\beta x^n dm(x) = \sum_{n=0}^\infty a_n \frac{\beta^{n+1} - \alpha^{n+1}}{n+1}$$

for all $-R < \alpha < \beta < R$. Indeed this follows from Corollary 7.27 since

$$\sum_{n=0}^\infty \int_\alpha^\beta |a_n|\, |x|^n\, dm(x) \le \sum_{n=0}^\infty \left(\int_0^{|\beta|} |a_n|\, |x|^n\, dm(x) + \int_0^{|\alpha|} |a_n|\, |x|^n\, dm(x) \right)$$
$$\le \sum_{n=0}^\infty |a_n|\, \frac{|\beta|^{n+1} + |\alpha|^{n+1}}{n+1} \le 2r \sum_{n=0}^\infty |a_n|\, r^n < \infty$$

where $r = \max(|\beta|, |\alpha|)$.

Example 7.42. Let $\{r_n\}_{n=1}^\infty$ be an enumeration of the points in $\mathbb{Q} \cap [0, 1]$ and define

$$f(x) = \sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}}$$

with the convention that

$$\frac{1}{\sqrt{|x - r_n|}} = 5 \text{ if } x = r_n.$$

Since, By Theorem 7.39,

$$\int_0^1 \frac{1}{\sqrt{|x - r_n|}}\, dx = \int_{r_n}^1 \frac{1}{\sqrt{x - r_n}} dx + \int_0^{r_n} \frac{1}{\sqrt{r_n - x}} dx$$
$$= 2\sqrt{x - r_n}|_{r_n}^1 - 2\sqrt{r_n - x}|_0^{r_n} = 2\left(\sqrt{1 - r_n} - \sqrt{r_n}\right)$$
$$\le 4,$$

we find

$$\int_{[0,1]} f(x) dm(x) = \sum_{n=1}^\infty 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x - r_n|}} dx \le \sum_{n=1}^\infty 2^{-n} 4 = 4 < \infty.$$

In particular, $m(f = \infty) = 0$, i.e. that $f < \infty$ for almost every $x \in [0, 1]$ and this implies that

$$\sum_{n=1}^\infty 2^{-n} \frac{1}{\sqrt{|x - r_n|}} < \infty \text{ for a.e. } x \in [0, 1].$$

This result is somewhat surprising since the singularities of the summands form a dense subset of $[0, 1]$.

Example 7.43. The following limit holds,

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n dm(x) = 1. \tag{7.19}$$

DCT Proof. To verify this, let $f_n(x) := \left(1 - \frac{x}{n}\right)^n 1_{[0,n]}(x)$. Then $\lim_{n\to\infty} f_n(x) = e^{-x}$ for all $x \ge 0$. Moreover by simple calculus[2] (or taking logarithms of Eq. (**??**) below) we have

$$\ln(1 - x) \le -x \text{ for } x < 1.$$

Therefore, for $x < n$, we have

$$\left(1 - \frac{x}{n}\right)^n = e^{n \ln\left(1 - \frac{x}{n}\right)} \le e^{-n\left(\frac{x}{n}\right)} = e^{-x}$$

from which it follows that

$$0 \le f_n(x) \le e^{-x} \text{ for all } x \ge 0.$$

From Example 7.40, we know

$$\int_0^\infty e^{-x} dm(x) = 1 < \infty,$$

[2] Indeed, $\ln(1 - x)$ is concave down and $y = -x$ is the tangent line to $y = \ln(1 - x)$ at $x = 0$.

so that e^{-x} is an integrable function on $[0,\infty)$. Hence by the dominated convergence theorem,

$$\lim_{n\to\infty}\int_0^n\left(1-\frac{x}{n}\right)^n dm(x)=\lim_{n\to\infty}\int_0^\infty f_n(x)dm(x)$$
$$=\int_0^\infty\lim_{n\to\infty}f_n(x)dm(x)=\int_0^\infty e^{-x}dm(x)=1.$$

MCT Proof. The limit in Eq. (7.19) may also be computed using the monotone convergence theorem. To do this we must show that $n\to f_n(x)$ is increasing in n for each x and for this it suffices to consider $n>x$. But for $n>x$,

$$\frac{d}{dn}\ln f_n(x)=\frac{d}{dn}\left[n\ln\left(1-\frac{x}{n}\right)\right]=\ln\left(1-\frac{x}{n}\right)+\frac{n}{1-\frac{x}{n}}\frac{x}{n^2}$$
$$=\ln\left(1-\frac{x}{n}\right)+\frac{\frac{x}{n}}{1-\frac{x}{n}}=h(x/n)$$

where, for $0\le y<1$,

$$h(y):=\ln(1-y)+\frac{y}{1-y}.$$

Since $h(0)=0$ and

$$h'(y)=-\frac{1}{1-y}+\frac{1}{1-y}+\frac{y}{(1-y)^2}>0$$

it follows that $h\ge 0$. Thus we have shown, $f_n(x)\uparrow e^{-x}$ as $n\to\infty$ as claimed.

Example 7.44 (Jordan's Lemma). In this example, let us consider the limit;

$$\lim_{n\to\infty}\int_0^\pi\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta.$$

Let

$$f_n(\theta):=1_{(0,\pi]}(\theta)\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}.$$

Then

$$|f_n|\le 1_{(0,\pi]}\in L^1(m)$$

and

$$\lim_{n\to\infty}f_n(\theta)=1_{(0,\pi]}(\theta)\,1_{\{\pi\}}(\theta)=1_{\{\pi\}}(\theta).$$

Therefore by the D.C.T.,

$$\lim_{n\to\infty}\int_0^\pi\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta=\int_{\mathbb{R}}1_{\{\pi\}}(\theta)\,dm(\theta)=m(\{\pi\})=0.$$

Example 7.45. Recall from Example 7.40 that

$$\lambda^{-1}=\int_{[0,\infty)}e^{-\lambda x}dm(x)\text{ for all }\lambda>0.$$

Let $\varepsilon>0$. For $\lambda\ge 2\varepsilon>0$ and $n\in\mathbb{N}$ there exists $C_n(\varepsilon)<\infty$ such that

$$0\le\left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x}=x^n e^{-\lambda x}\le C(\varepsilon)e^{-\varepsilon x}.$$

Using this fact, Corollary 7.29 and induction gives

$$n!\lambda^{-n-1}=\left(-\frac{d}{d\lambda}\right)^n\lambda^{-1}=\int_{[0,\infty)}\left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x}dm(x)$$
$$=\int_{[0,\infty)}x^n e^{-\lambda x}dm(x).$$

That is

$$n!=\lambda^n\int_{[0,\infty)}x^n e^{-\lambda x}dm(x). \tag{7.20}$$

Remark 7.46. Corollary 7.29 may be generalized by allowing the hypothesis to hold for $x\in X\setminus E$ where $E\in\mathcal{B}$ is a **fixed** null set, i.e. E must be independent of t. Consider what happens if we formally apply Corollary 7.29 to $g(t):=\int_0^\infty 1_{x\le t}dm(x)$,

$$\dot g(t)=\frac{d}{dt}\int_0^\infty 1_{x\le t}dm(x)\overset{?}{=}\int_0^\infty\frac{\partial}{\partial t}1_{x\le t}dm(x).$$

The last integral is zero since $\frac{\partial}{\partial t}1_{x\le t}=0$ unless $t=x$ in which case it is not defined. On the other hand $g(t)=t$ so that $\dot g(t)=1$. (The reader should decide which hypothesis of Corollary 7.29 has been violated in this example.)

Exercise 7.6 (Folland 2.28 on p. 60.). Compute the following limits and justify your calculations:

1. $\lim_{n\to\infty}\int_0^\infty\frac{\sin(\frac{x}{n})}{(1+\frac{x}{n})^n}\,dx.$
2. $\lim_{n\to\infty}\int_0^1\frac{1+nx^2}{(1+x^2)^n}\,dx$
3. $\lim_{n\to\infty}\int_0^\infty\frac{n\sin(x/n)}{x(1+x^2)}\,dx$
4. For all $a\in\mathbb{R}$ compute,

$$f(a):=\lim_{n\to\infty}\int_a^\infty n(1+n^2x^2)^{-1}dx.$$

Exercise 7.7 (Integration by Parts). Suppose that $f, g : \mathbb{R} \to \mathbb{R}$ are two continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $\mathbb{R}$. Prove the following integration by parts formula;

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\, dx = -\int_{\mathbb{R}} f(x) \cdot g'(x)\, dx. \tag{7.21}$$

Similarly show that if Suppose that $f, g : [0,\infty) \to [0,\infty)$ are two continuously differentiable functions such that $f'g$, fg', and fg are all Lebesgue integrable functions on $[0,\infty)$, then

$$\int_0^\infty f'(x) \cdot g(x)\, dx = -f(0)\, g(0) - \int_0^\infty f(x) \cdot g'(x)\, dx. \tag{7.22}$$

Outline: 1. First notice that Eq. (7.21) holds if $f(x) = 0$ for $|x| \geq N$ for some $N < \infty$ by undergraduate calculus.

2. Let $\psi : \mathbb{R} \to [0,1]$ be a continuously differentiable function such that $\psi(x) = 1$ if $|x| \leq 1$ and $\psi(x) = 0$ if $|x| \geq 2$. For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ Write out the identity in Eq. (7.21) with $f(x)$ being replaced by $f(x)\, \psi_\varepsilon(x)$.

3. Now use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in the identity you found in step 2.

4. A similar outline works to prove Eq. (7.22).

Solution to Exercise (7.7). If f has compact support in $[-N, N]$ for some $N < \infty$, then by undergraduate integration by parts,

$$\begin{aligned}\int_{\mathbb{R}} f'(x) \cdot g(x)\, dx &= \int_{-N}^{N} f'(x) \cdot g(x)\, dx \\ &= f(x)\, g(x)\, |_{-N}^{N} - \int_{-N}^{N} f(x) \cdot g'(x)\, dx \\ &= -\int_{-N}^{N} f(x) \cdot g'(x)\, dx = -\int_{\mathbb{R}} f(x) \cdot g'(x)\, dx.\end{aligned}$$

Similarly if f has compact support in $[0,\infty)$, then

$$\begin{aligned}\int_0^\infty f'(x) \cdot g(x)\, dx &= \int_0^N f'(x) \cdot g(x)\, dx \\ &= f(x)\, g(x)\, |_0^N - \int_0^N f(x) \cdot g'(x)\, dx \\ &= -f(0)\, g(0) - \int_0^N f(x) \cdot g'(x)\, dx \\ &= -f(0) - \int_0^\infty f(x) \cdot g'(x)\, dx.\end{aligned}$$

For general f we may apply this identity with $f(x)$ replaced by $\psi_\varepsilon(x)\, f(x)$ to learn,

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\, \psi_\varepsilon(x)\, dx + \int_{\mathbb{R}} f(x) \cdot g(x)\, \psi'_\varepsilon(x)\, dx = -\int_{\mathbb{R}} \psi_\varepsilon(x)\, f(x) \cdot g'(x)\, dx. \tag{7.23}$$

Since $\psi_\varepsilon(x) \to 1$ boundedly and $|\psi'_\varepsilon(x)| = \varepsilon\, |\psi'(\varepsilon x)| \leq C\varepsilon$, we may use the DCT to conclude,

$$\begin{aligned}&\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} f'(x) \cdot g(x)\, \psi_\varepsilon(x)\, dx = \int_{\mathbb{R}} f'(x) \cdot g(x)\, dx, \\ &\lim_{\varepsilon \downarrow 0} \int_{\mathbb{R}} f(x) \cdot g'(x)\, \psi_\varepsilon(x)\, dx = \int_{\mathbb{R}} f(x) \cdot g'(x)\, dx, \text{ and} \\ &\left| \int_{\mathbb{R}} f(x) \cdot g(x)\, \psi'_\varepsilon(x)\, dx \right| \leq C\varepsilon \cdot \int_{\mathbb{R}} |f(x) \cdot g(x)|\, dx \to 0 \text{ as } \varepsilon \downarrow 0.\end{aligned}$$

Therefore passing to the limit as $\varepsilon \downarrow 0$ in Eq. (7.23) completes the proof of Eq. (7.21). Equation (7.22) is proved in the same way.

Definition 7.47 (Gamma Function). *The* ***Gamma function***, $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ *is defined by*

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \tag{7.24}$$

(The reader should check that $\Gamma(x) < \infty$ for all $x > 0$.)

Here are some of the more basic properties of this function.

Example 7.48 (Γ – function properties). Let Γ be the gamma function, then;

1. $\Gamma(1) = 1$ as is easily verified.
2. $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$ as follows by integration by parts;

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty e^{-u}\; u^{x+1}\, \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\, e^{-u} \right) du \\ &= x \int_0^\infty u^{x-1}\; e^{-u}\; du = x\; \Gamma(x).\end{aligned}$$

In particular, it follows from items 1. and 2. and induction that

$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{7.25}$$

(Equation 7.25was also proved in Eq. (7.20).)

3. $\Gamma(1/2)=\sqrt{\pi}$. This last assertion is a bit trickier. One proof is to make use of the fact (proved below in Lemma 9.27) that
$$\int_{-\infty}^{\infty} e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a>0. \tag{7.26}$$
Taking $a=1$ and making the change of variables, $u=r^2$ below implies,
$$\sqrt{\pi}=\int_{-\infty}^{\infty} e^{-r^2}dr = 2\int_0^\infty u^{-1/2}e^{-u}du = \Gamma(1/2).$$
$$\Gamma(1/2) = 2\int_0^\infty e^{-r^2}dr = \int_{-\infty}^{\infty} e^{-r^2}dr$$
$$= I_1(1) = \sqrt{\pi}.$$
4. A simple induction argument using items 2. and 3. now shows that
$$\Gamma\left(n+\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi}$$
where $(-1)!! := 1$ and $(2n-1)!! = (2n-1)(2n-3)\dots 3\cdot 1$ for $n\in\mathbb{N}$.

7.4 Densities and Change of Variables Theorems

Exercise 7.8. Let $(X,\mathcal{M},\mu)$ be a measure space and $\rho : X\to[0,\infty]$ be a measurable function. For $A\in\mathcal{M}$, set $\nu(A) := \int_A \rho d\mu$.

1. Show $\nu : \mathcal{M}\to[0,\infty]$ is a measure.
2. Let $f : X\to[0,\infty]$ be a measurable function, show
$$\int_X f d\nu = \int_X f\rho d\mu. \tag{7.27}$$
Hint: first prove the relationship for characteristic functions, then for simple functions, and then for general positive measurable functions.
3. Show that a measurable function $f : X\to\mathbb{C}$ is in $L^1(\nu)$ iff $|f|\rho\in L^1(\mu)$ and if $f\in L^1(\nu)$ then Eq. (7.27) still holds.

Solution to Exercise (7.8). The fact that ν is a measure follows easily from Corollary 7.6. Clearly Eq. (7.27) holds when $f=1_A$ by definition of ν. It then holds for positive simple functions, f, by linearity. Finally for general $f\in L^+$, choose simple functions, φ_n, such that $0\le\varphi_n\uparrow f$. Then using MCT twice we find
$$\int_X f d\nu = \lim_{n\to\infty}\int_X \varphi_n d\nu = \lim_{n\to\infty}\int_X \varphi_n\rho d\mu = \int_X \lim_{n\to\infty}\varphi_n\rho d\mu = \int_X f\rho d\mu.$$
By what we have just proved, for all $f : X\to\mathbb{C}$ we have
$$\int_X |f|\, d\nu = \int_X |f|\,\rho d\mu$$
so that $f\in L^1(\mu)$ iff $|f|\rho\in L^1(\mu)$. If $f\in L^1(\mu)$ and f is real,
$$\int_X f d\nu = \int_X f_+ d\nu - \int_X f_- d\nu = \int_X f_+\rho d\mu - \int_X f_-\rho d\mu$$
$$= \int_X [f_+\rho - f_-\rho]\, d\mu = \int_X f\rho d\mu.$$
The complex case easily follows from this identity.

Notation 7.49 *It is customary to informally describe ν defined in Exercise 7.8 by writing $d\nu=\rho d\mu$.*

Exercise 7.9. Let $(X,\mathcal{M},\mu)$ be a measure space, $(Y,\mathcal{F})$ be a measurable space and $f : X\to Y$ be a measurable map. Recall that $\nu=f_*\mu : \mathcal{F}\to[0,\infty]$ defined by $\nu(A) := \mu(f^{-1}(A))$ for all $A\in\mathcal{F}$ is a measure on $\mathcal{F}$.

1. Show
$$\int_Y g d\nu = \int_X (g\circ f)\, d\mu \tag{7.28}$$
for all measurable functions $g : Y\to[0,\infty]$. **Hint:** see the hint from Exercise 7.8.
2. Show a measurable function $g : Y\to\mathbb{C}$ is in $L^1(\nu)$ iff $g\circ f\in L^1(\mu)$ and that Eq. (7.28) holds for all $g\in \mathrm{L}^1(\nu)$.

Remark 7.50. If X is a random variable on a probability space, $(\Omega,\mathcal{B},P)$, and $F(x) := P(X\le x)$. Then
$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\, dF(x) \tag{7.29}$$
where $dF(x)$ is shorthand for $d\mu_F(x)$ and μ_F is the unique probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\mu_F((-\infty,x]) = F(x)$ for all $x\in\mathbb{R}$. Moreover if $F:\mathbb{R}\to[0,1]$ happens to be C^1-function, then
$$d\mu_F(x) = F'(x)\, dm(x) \tag{7.30}$$
and Eq. (7.29) may be written as

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x)\, dm(x). \tag{7.31}$$

To verify Eq. (7.30) it suffices to observe, by the fundamental theorem of calculus, that

$$\mu_F((a,b]) = F(b) - F(a) = \int_a^b F'(x)\, dx = \int_{(a,b]} F' dm.$$

From this equation we may deduce that $\mu_F(A) = \int_A F' dm$ for all $A \in \mathcal{B}_{\mathbb{R}}$.

Exercise 7.10. Let $F : \mathbb{R} \to \mathbb{R}$ be a C^1-function such that $F'(x) > 0$ for all $x \in \mathbb{R}$ and $\lim_{x\to\pm\infty} F(x) = \pm\infty$. (Notice that F is strictly increasing so that $F^{-1} : \mathbb{R} \to \mathbb{R}$ exists and moreover, by the inverse function theorem that F^{-1} is a C^1 – function.) Let m be Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$ and

$$\nu(A) = m(F(A)) = m(\left(F^{-1}\right)^{-1}(A)) = \left(F_*^{-1} m\right)(A)$$

for all $A \in \mathcal{B}_{\mathbb{R}}$. Show $d\nu = F' dm$. Use this result to prove the change of variable formula,

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h dm \tag{7.32}$$

which is valid for all Borel measurable functions $h : \mathbb{R} \to [0,\infty]$.

Hint: Start by showing $d\nu = F' dm$ on sets of the form $A = (a,b]$ with $a, b \in \mathbb{R}$ and $a < b$. Then use the uniqueness assertions in Exercise 5.11 to conclude $d\nu = F' dm$ on all of $\mathcal{B}_{\mathbb{R}}$. To prove Eq. (7.32) apply Exercise 7.9 with $g = h \circ F$ and $f = F^{-1}$.

Solution to Exercise (7.10). Let $d\mu = F' dm$ and $A = (a,b]$, then

$$\nu((a,b]) = m(F((a,b])) = m((F(a), F(b)]) = F(b) - F(a)$$

while

$$\mu((a,b]) = \int_{(a,b]} F' dm = \int_a^b F'(x) dx = F(b) - F(a).$$

It follows that both $\mu = \nu = \mu_F$ – where μ_F is the measure described in Theorem 5.33. By Exercise 7.9 with $g = h \circ F$ and $f = F^{-1}$, we find

$$\begin{aligned}\int_{\mathbb{R}} h \circ F \cdot F' dm &= \int_{\mathbb{R}} h \circ F d\nu = \int_{\mathbb{R}} h \circ F d\left(F_*^{-1} m\right) = \int_{\mathbb{R}} (h \circ F) \circ F^{-1} dm \\ &= \int_{\mathbb{R}} h dm.\end{aligned}$$

This result is also valid for all $h \in L^1(m)$.

7.5 Some Common Continuous Distributions

Example 7.51 (Uniform Distribution). Suppose that X has the uniform distribution in $[0,b]$ for some $b \in (0,\infty)$, i.e. $X_* P = \frac{1}{b} \cdot m$ on $[0,b]$. More explicitly,

$$\mathbb{E}[f(X)] = \frac{1}{b}\int_0^b f(x)\, dx \text{ for all bounded measurable } f.$$

The moment generating function for X is;

$$\begin{aligned} M_X(t) &= \frac{1}{b}\int_0^b e^{tx} dx = \frac{1}{bt}\left(e^{tb} - 1\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{n!}(bt)^{n-1} = \sum_{n=0}^{\infty} \frac{b^n}{(n+1)!} t^n. \end{aligned}$$

On the other hand (see Proposition 7.32),

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n.$$

Thus it follows that

$$\mathbb{E}X^n = \frac{b^n}{n+1}.$$

Of course this may be calculated directly just as easily,

$$\mathbb{E}X^n = \frac{1}{b}\int_0^b x^n dx = \frac{1}{b(n+1)} x^{n+1}|_0^b = \frac{b^n}{n+1}.$$

Definition 7.52. *A random variable $T \ge 0$ is said to be* ***exponential with parameter*** *$\lambda \in [0,\infty)$ provided, $P(T > t) = e^{-\lambda t}$ for all $t \ge 0$. We will write $T \stackrel{d}{=} E(\lambda)$ for short.*

If $\lambda > 0$, we have

$$P(T > t) = e^{-\lambda t} = \int_t^\infty \lambda e^{-\lambda \tau} d\tau$$

from which it follows that $P(T \in (t, t+dt)) = \lambda 1_{t \ge 0} e^{-\lambda t} dt$. Applying Corollary 7.29 repeatedly implies,

$$\mathbb{E}T = \int_0^\infty \tau \lambda e^{-\lambda \tau} d\tau = \lambda \left(-\frac{d}{d\lambda}\right) \int_0^\infty e^{-\lambda \tau} d\tau = \lambda \left(-\frac{d}{d\lambda}\right) \lambda^{-1} = \lambda^{-1}$$

and more generally that

$$\mathbb{E}T^k = \int_0^\infty \tau^k e^{-\lambda\tau}\lambda d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k \int_0^\infty e^{-\lambda\tau}d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k \lambda^{-1} = k!\lambda^{-k}. \tag{7.33}$$

In particular we see that

$$\mathrm{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \tag{7.34}$$

Alternatively we may compute the moment generating function for T,

$$\begin{aligned} M_T(a) &:= \mathbb{E}\left[e^{aT}\right] = \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau}d\tau \\ &= \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau}d\tau = \frac{\lambda}{\lambda - a} = \frac{1}{1 - a\lambda^{-1}} \end{aligned} \tag{7.35}$$

which is valid for $a < \lambda$. On the other hand (see Proposition 7.32), we know that

$$\mathbb{E}\left[e^{aT}\right] = \sum_{n=0}^\infty \frac{a^n}{n!}\mathbb{E}[T^n] \text{ for } |a| < \lambda.$$

Comparing this with Eq. (7.35) again shows that Eq. (7.33) is valid.

Here is yet another way to understand and generalize Eq. (7.35). We simply make the change of variables, $u = \lambda\tau$ in the integral in Eq. (7.33) to learn,

$$\mathbb{E}T^k = \lambda^{-k}\int_0^\infty u^k e^{-u}d\tau = \lambda^{-k}\Gamma(k+1).$$

This last equation is valid for all $k \in (-1,\infty)$ – in particular k need not be an integer.

Theorem 7.53 (Memoryless property). *A random variable, $T \in (0,\infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$P(T > s+t | T > s) = P(T > t) \text{ for all } s,t \ge 0,$$

where as usual, $P(A|B) := P(A\cap B)/P(B)$ when $p(B) > 0$. (Note that $T \overset{d}{=} E(0)$ means that $P(T > t) = e^{0t} = 1$ for all $t > 0$ and therefore that $T = \infty$ a.s.)

Proof. (The following proof is taken from [33].) Suppose first that $T \overset{d}{=} E(\lambda)$ for some $\lambda > 0$. Then

$$P(T > s+t | T > s) = \frac{P(T > s+t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t).$$

For the converse, let $g(t) := P(T > t)$, then by assumption,

$$\frac{g(t+s)}{g(s)} = P(T > s+t | T > s) = P(T > t) = g(t)$$

whenever $g(s) \neq 0$ and $g(t)$ is a decreasing function. Therefore if $g(s) = 0$ for some $s > 0$ then $g(t) = 0$ for all $t > s$. Thus it follows that

$$g(t+s) = g(t)\,g(s) \text{ for all } s,t \ge 0.$$

Since $T > 0$, we know that $g(1/n) = P(T > 1/n) > 0$ for some n and therefore, $g(1) = g(1/n)^n > 0$ and we may write $g(1) = e^{-\lambda}$ for some $0 \le \lambda < \infty$.

Observe for $p,q \in \mathbb{N}$, $g(p/q) = g(1/q)^p$ and taking $p = q$ then shows, $e^{-\lambda} = g(1) = g(1/q)^q$. Therefore, $g(p/q) = e^{-\lambda p/q}$ so that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{Q}_+ := \mathbb{Q}\cap\mathbb{R}_+$. Given $r,s \in \mathbb{Q}_+$ and $t \in \mathbb{R}$ such that $r \le t \le s$ we have, since g is decreasing, that

$$e^{-\lambda r} = g(r) \ge g(t) \ge g(s) = e^{-\lambda s}.$$

Hence letting $s \uparrow t$ and $r \downarrow t$ in the above equations shows that $g(t) = e^{-\lambda t}$ for all $t \in \mathbb{R}_+$ and therefore $T \overset{d}{=} E(\lambda)$. ■

Exercise 7.11 (Gamma Distributions). Let X be a positive random variable. For $k,\theta > 0$, we say that $X \overset{d}{=}\mathrm{Gamma}(k,\theta)$ if

$$(X_*P)(dx) = f(x;k,\theta)\,dx \text{ for } x > 0,$$

where

$$f(x;k,\theta) := x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)} \text{ for } x > 0, \text{ and } k,\theta > 0.$$

Find the moment generating function (see Definition 7.31), $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ for $t < \theta^{-1}$. Differentiate your result in t to show

$$\mathbb{E}[X^m] = k(k+1)\dots(k+m-1)\,\theta^m \text{ for all } m \in \mathbb{N}_0.$$

In particular, $\mathbb{E}[X] = k\theta$ and $\mathrm{Var}(X) = k\theta^2$. (Notice that when $k = 1$ and $\theta = \lambda^{-1}$, $X \overset{d}{=} E(\lambda)$.)

7.5.1 Normal (Gaussian) Random Variables

Definition 7.54 (Normal / Gaussian Random Variables). *A random variable, Y,* ***is normal with mean*** *μ* ***standard deviation*** *σ^2 iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2}dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{7.36}$$

We will abbreviate this by writing $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$. *When* $\mu = 0$ *and* $\sigma^2 = 1$ *we will simply write* N *for* $N(0,1)$ *and if* $Y \overset{d}{=} N$, *we will say* Y *is a* ***standard normal*** *random variable.*

Observe that Eq. (7.36) is equivalent to writing

$$\mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y)\, e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy$$

for all bounded measurable functions, $f : \mathbb{R} \to \mathbb{R}$. Also observe that $Y \overset{d}{=} N\left(\mu, \sigma^2\right)$ is equivalent to $Y \overset{d}{=} \sigma N + \mu$. Indeed, by making the change of variable, $y = \sigma x + \mu$, we find

$$\begin{aligned}\mathbb{E}[f(\sigma N + \mu)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\sigma x + \mu)\, e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(y)\, e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y)\, e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy.\end{aligned}$$

Lastly the constant, $\left(2\pi\sigma^2\right)^{-1/2}$ is chosen so that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}y^2} dy = 1,$$

see Example 7.48 and Lemma 9.27.

Exercise 7.12. Suppose that $X \overset{d}{=} N(0,1)$ and $f : \mathbb{R} \to \mathbb{R}$ is a C^1 – function such that $Xf(X)$, $f'(X)$ and $f(X)$ are all integrable random variables. Show

$$\mathbb{E}[Xf(X)] = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \frac{d}{dx} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(x)\, e^{-\frac{1}{2}x^2} dx = \mathbb{E}[f'(X)].$$

Example 7.55. Suppose that $X \overset{d}{=} N(0,1)$ and define $\alpha_k := \mathbb{E}\left[X^{2k}\right]$ for all $k \in \mathbb{N}_0$. By Exercise 7.12,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1} \cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

Hence it follows that

$$\alpha_1 = \alpha_0 = 1,\ \alpha_2 = 3\alpha_1 = 3,\ \alpha_3 = 5 \cdot 3$$

and by a simple induction argument,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!, \tag{7.37}$$

where $(-1)!! := 0$. Actually we can use the Γ – function to say more. Namely for any $\beta > -1$,

$$\mathbb{E}|X|^\beta = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^\beta e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x^\beta e^{-\frac{1}{2}x^2} dx.$$

Now make the change of variables, $y = x^2/2$ (i.e. $x = \sqrt{2y}$ and $dx = \frac{1}{\sqrt{2}} y^{-1/2} dy$) to learn,

$$\begin{aligned}\mathbb{E}|X|^\beta &= \frac{1}{\sqrt{\pi}} \int_0^\infty (2y)^{\beta/2} e^{-y} y^{-1/2} dy \\ &= \frac{1}{\sqrt{\pi}} 2^{\beta/2} \int_0^\infty y^{(\beta+1)/2} e^{-y} y^{-1} dy = \frac{1}{\sqrt{\pi}} 2^{\beta/2} \Gamma\left(\frac{\beta+1}{2}\right).\end{aligned}$$

Exercise 7.13. Suppose that $X \overset{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = \exp\left(-\lambda^2/2\right). \tag{7.38}$$

Hint: Use Corollary 7.29 to show, $f'(\lambda) = i\mathbb{E}\left[Xe^{i\lambda X}\right]$ and then use Exercise 7.12 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Solution to Exercise (7.13). Using Corollary 7.29 and Exercise 7.12,

$$\begin{aligned}f'(\lambda) &= i\mathbb{E}\left[Xe^{i\lambda X}\right] = i\mathbb{E}\left[\frac{d}{dX} e^{i\lambda X}\right] \\ &= i \cdot (i\lambda)\, \mathbb{E}\left[e^{i\lambda X}\right] = -\lambda f(\lambda) \text{ with } f(0) = 1.\end{aligned}$$

Solving for the unique solution of this differential equation gives Eq. (7.38).

Exercise 7.14. Suppose that $X \overset{d}{=} N(0,1)$ and $t \in \mathbb{R}$. Show $\mathbb{E}\left[e^{tX}\right] = \exp\left(t^2/2\right)$. (You could follow the hint in Exercise 7.13 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 7.15. Use Exercise 7.14 and Proposition 7.32 to give another proof that $\mathbb{E}X^{2k} = (2k-1)!!$ when $X \overset{d}{=} N(0,1)$.

Exercise 7.16. Let $X \overset{d}{=} N(0,1)$ and $\alpha \in \mathbb{R}$, find $\rho : \mathbb{R}_+ \to \mathbb{R}_+ := (0,\infty)$ such that

$$\mathbb{E}\left[f\left(|X|^\alpha\right)\right] = \int_{\mathbb{R}_+} f(x)\,\rho(x)\,dx$$

for all continuous functions, $f : \mathbb{R}_+ \to \mathbb{R}$ with compact support in $\mathbb{R}_+$.

Lemma 7.56 (Gaussian tail estimates). *Suppose that X is a standard normal random variable, i.e.*

$$P(X \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}},$$

then for all $x \geq 0$,

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right) \leq \frac{1}{2} e^{-x^2/2}. \tag{7.39}$$

Moreover (see [35, Lemma 2.5]),

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) \tag{7.40}$$

which combined with Eq. (7.39) proves Mill's ratio (see [15]);

$$\lim_{x\to\infty} \frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x} e^{-x^2/2}} = 1. \tag{7.41}$$

Proof. See Figure 7.1 where; the green curve is the plot of $P(X \geq x)$, the black is the plot of

$$\min\left(\frac{1}{2} - \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}\right),$$

the red is the plot of $\frac{1}{2} e^{-x^2/2}$, and the blue is the plot of

$$\max\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right).$$

The formal proof of these estimates for the reader who is not convinced by Figure 7.1 is given below.

We begin by observing that

$$P(X \geq x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{y}{x} e^{-y^2/2} dy$$
$$\leq -\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-y^2/2}\Big|_x^{-\infty} = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}.$$

On the other hand we have,

$$P(X \geq x) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy \tag{7.42}$$
$$\leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-x^2/2} dy \leq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x.$$

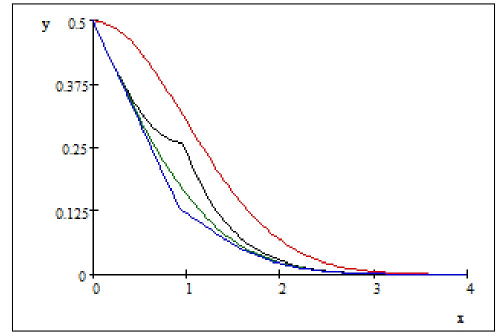


Fig. 7.1. Plots of $P(X \geq x)$ and its estimates.

The last two equations give the first equality in Eq. (7.39). To prove the second equality observe that $\sqrt{2\pi} > 2$, so $\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2}$ if $x \geq 1$. For $x \leq 1$ we must show,

$$\frac{1}{2} - \frac{x}{\sqrt{2\pi}} e^{-x^2/2} \leq \frac{1}{2} e^{-x^2/2}$$

or equivalently that $f(x) := e^{x^2/2} - \sqrt{\frac{2}{\pi}} x \leq 1$ for $0 \leq x \leq 1$. Since f is convex $\left(f''(x) = (x^2+1) e^{x^2/2} > 0\right)$, $f(0) = 1$ and $f(1) \cong 0.85 < 1$, it follows that $f \leq 1$ on $[0,1]$. This proves the second inequality in Eq. (7.39).

It follows from Eq. (7.42) that

$$P(X \geq x) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy$$
$$\geq \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x 1 dy = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} x \text{ for all } x \geq 0.$$

So to finish the proof of Eq. (7.40) we must show,

$$f(x) := \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} - (1+x^2) P(X \geq x)$$
$$= \frac{1}{\sqrt{2\pi}} \left[x e^{-x^2/2} - (1+x^2) \int_x^\infty e^{-y^2/2} dy\right] \leq 0 \text{ for all } 0 \leq x < \infty.$$

This follows by observing that $f(0) = -1/2 < 0$, $\lim_{x\uparrow\infty} f(x) = 0$ and

$$f'(x) = \frac{1}{\sqrt{2\pi}} \left[e^{-x^2/2}(1-x^2) - 2xP(X \geq x) + (1+x^2) e^{-x^2/2}\right]$$
$$= 2\left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2} - xP(X \geq y)\right) \geq 0,$$

where the last inequality is a consequence Eq. (7.39). ■

7.6 Comparison of the Lebesgue and the Riemann Integral*

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a,b] \to \mathbb{R}$ be a bounded function. A partition of $[a,b]$ is a finite subset $\pi \subset [a,b]$ containing $\{a,b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{7.43}$$

of $[a,b]$ let

$$\text{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1,\dots,n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \quad m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1},t_j]}, \quad g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1},t_j]} \text{ and}$$

$$S_\pi f = \sum M_j(t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j(t_j - t_{j-1}).$$

Notice that

$$S_\pi f = \int_a^b G_\pi dm \text{ and } s_\pi f = \int_a^b g_\pi dm.$$

The upper and lower Riemann integrals are defined respectively by

$$\overline{\int_a^b} f(x)dx = \inf_\pi S_\pi f \text{ and } \underline{\int_b^a} f(x)dx = \sup_\pi \; s_\pi f.$$

Definition 7.57. *The function f is **Riemann integrable** iff $\overline{\int_a^b} f = \underline{\int_a^b} f \in \mathbb{R}$ and which case the Riemann integral $\int_a^b f$ is defined to be the common value:*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

The proof of the following Lemma is left to the reader as Exercise 7.27.

Lemma 7.58. *If π' and π are two partitions of $[a,b]$ and $\pi \subset \pi'$ then*

$$G_\pi \ge G_{\pi'} \ge f \ge g_{\pi'} \ge g_\pi \text{ and}$$
$$S_\pi f \ge S_{\pi'} f \ge s_{\pi'} f \ge s_\pi f.$$

There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that mesh$(\pi_k) \downarrow 0$ *and*

$$S_{\pi_k} f \downarrow \overline{\int_a^b} f \text{ and } s_{\pi_k} f \uparrow \underline{\int_a^b} f \text{ as } k \to \infty.$$

If we let

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{7.44}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} gdm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{7.45}$$

and

$$\int_{[a,b]} Gdm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{7.46}$$

Notation 7.59 *For $x \in [a,b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \; \sup\{f(y) : |y-x| \le \varepsilon, \; y \in [a,b]\} \text{ and}$$
$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \; \inf\{f(y) : |y-x| \le \varepsilon, \; y \in [a,b]\}.$$

Lemma 7.60. *The functions $H, h : [a,b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \le f(x) \le H(x)$ for all $x \in [a,b]$ and $h(x) = H(x)$ iff f is continuous at x.*
2. *If $\{\pi_k\}_{k=1}^\infty$ is any increasing sequence of partitions such that* mesh$(\pi_k) \downarrow 0$ *and G and g are defined as in Eq. (7.44), then*
$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall \; x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{7.47}$$
(Note π is a countable set.)
3. *H and h are Borel measurable.*

Proof. Let $G_k := G_{\pi_k} \downarrow G$ and $g_k := g_{\pi_k} \uparrow g$.

1. It is clear that $h(x) \le f(x) \le H(x)$ for all x and $H(x) = h(x)$ iff $\lim_{y\to x} f(y)$ exists and is equal to $f(x)$. That is $H(x) = h(x)$ iff f is continuous at x.
2. For $x \notin \pi$,
$$G_k(x) \ge H(x) \ge f(x) \ge h(x) \ge g_k(x) \; \forall \; k$$
and letting $k \to \infty$ in this equation implies
$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x) \; \forall \; x \notin \pi. \tag{7.48}$$
Moreover, given $\varepsilon > 0$ and $x \notin \pi$,
$$\sup\{f(y) : |y-x| \le \varepsilon, \; y \in [a,b]\} \ge G_k(x)$$

for all k large enough, since eventually $G_k(x)$ is the supremum of $f(y)$ over some interval contained in $[x-\varepsilon, x+\varepsilon]$. Again letting $k \to \infty$ implies $\sup_{|y-x|\le\varepsilon} f(y) \ge G(x)$ and therefore, that

$$H(x) = \limsup_{y\to x} f(y) \ge G(x)$$

for all $x \notin \pi$. Combining this equation with Eq. (7.48) then implies $H(x) = G(x)$ if $x \notin \pi$. A similar argument shows that $h(x) = g(x)$ if $x \notin \pi$ and hence Eq. (7.47) is proved.

3. The functions G and g are limits of measurable functions and hence measurable. Since $H = G$ and $h = g$ except possibly on the countable set π, both H and h are also Borel measurable. (You justify this statement.)

■

Theorem 7.61. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{7.49}$$

and the following statements are equivalent:

1. *$H(x) = h(x)$ for m -a.e. x,*
2. *the set*
$$E := \{x \in [a,b] : f \text{ is discontinuous at } x\}$$
is an $\bar{m}$ – null set.
3. *f is Riemann integrable.*

If f is Riemann integrable then f is Lebesgue measurable[3]*, i.e. f is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue σ – algebra and $\mathcal{B}$ is the Borel σ – algebra on $[a,b]$. Moreover if we let $\bar{m}$ denote the completion of m, then*

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{7.50}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 7.58 and let G and g be defined as in Lemma 7.60. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (7.49) is a consequence of Eqs. (7.45) and (7.46). From Eq. (7.49), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \le f \le H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (7.47), the remaining assertions including Eq. (7.50) are now consequences of Lemma 7.64. ■

Notation 7.62 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

7.7 Measurability on Complete Measure Spaces*

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

Proposition 7.63. *Suppose that $(X, \mathcal{B}, \mu)$ is a complete measure space*[4] *and $f : X \to \mathbb{R}$ is measurable.*

1. *If $g : X \to \mathbb{R}$ is a function such that $f(x) = g(x)$ for μ – a.e. x, then g is measurable.*
2. *If $f_n : X \to \mathbb{R}$ are measurable and $f : X \to \mathbb{R}$ is a function such that $\lim_{n\to\infty} f_n = f$, μ - a.e., then f is measurable as well.*

Proof. 1. Let $E = \{x : f(x) \neq g(x)\}$ which is assumed to be in $\mathcal{B}$ and $\mu(E) = 0$. Then $g = 1_{E^c} f + 1_E g$ since $f = g$ on E^c. Now $1_{E^c} f$ is measurable so g will be measurable if we show $1_E g$ is measurable. For this consider,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{7.51}$$

Since $(1_E g)^{-1}(B) \subset E$ if $0 \notin B$ and $\mu(E) = 0$, it follow by completeness of $\mathcal{B}$ that $(1_E g)^{-1}(B) \in \mathcal{B}$ if $0 \notin B$. Therefore Eq. (7.51) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ■

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0,1,2\}$, $\mathcal{B} = \{\{0\}, \{1,2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

Lemma 7.64. *Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\bar{\mathcal{M}}$ is the completion of $\mathcal{M}$ relative to μ and $\bar{\mu}$ is the extension of μ to $\bar{\mathcal{M}}$. Then a function $f : X \to \mathbb{R}$ is $(\bar{\mathcal{M}}, \mathcal{B} = \mathcal{B}_{\mathbb{R}})$ – measurable iff there exists a function $g : X \to \mathbb{R}$*

[3] f need not be Borel measurable.

[4] Recall this means that if $N \subset X$ is a set such that $N \subset A \in \mathcal{M}$ and $\mu(A) = 0$, then $N \in \mathcal{M}$ as well.

that is $(\mathcal{M}, \mathcal{B})$ – measurable such $E = \{x : f(x) \neq g(x)\} \in \bar{\mathcal{M}}$ and $\bar{\mu}(E) = 0$, i.e. $f(x) = g(x)$ for $\bar{\mu}$ – a.e. x. Moreover for such a pair f and g, $f \in L^1(\bar{\mu})$ iff $g \in L^1(\mu)$ and in which case

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

Proof. Suppose first that such a function g exists so that $\bar{\mu}(E) = 0$. Since g is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 7.63 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (7.52) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{7.52}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (7.52) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 6.39 it holds for all $\mathcal{M}$ – measurable functions $g : X \to [0, \infty]$. The rest of the assertions follow in the standard way by considering $(\operatorname{Re} g)_\pm$ and $(\operatorname{Im} g)_\pm$. ∎

7.8 More Exercises

Exercise 7.17. Let μ be a measure on an algebra $\mathcal{A} \subset 2^X$, then $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$ for all $A, B \in \mathcal{A}$.

Exercise 7.18 (From problem 12 on p. 27 of Folland.). Let $(X, \mathcal{M}, \mu)$ be a finite measure space and for $A, B \in \mathcal{M}$ let $\rho(A, B) = \mu(A \Delta B)$ where $A \Delta B = (A \setminus B) \cup (B \setminus A)$. It is clear that $\rho(A, B) = \rho(B, A)$. Show:

1. ρ satisfies the triangle inequality:
$$\rho(A, C) \leq \rho(A, B) + \rho(B, C) \text{ for all } A, B, C \in \mathcal{M}.$$
2. Define $A \sim B$ iff $\mu(A \Delta B) = 0$ and notice that $\rho(A, B) = 0$ iff $A \sim B$. Show "$\sim$ " is an equivalence relation.
3. Let $\mathcal{M}/\sim$ denote $\mathcal{M}$ modulo the equivalence relation, $\sim$, and let $[A] := \{B \in \mathcal{M} : B \sim A\}$. Show that $\bar{\rho}([A], [B]) := \rho(A, B)$ is gives a well defined metric on $\mathcal{M}/\sim$.
4. Similarly show $\tilde{\mu}([A]) = \mu(A)$ is a well defined function on $\mathcal{M}/\sim$ and show $\tilde{\mu} : (\mathcal{M}/\sim) \to \mathbb{R}_+$ is $\bar{\rho}$ – continuous.

Exercise 7.19. Suppose that $\mu_n : \mathcal{M} \to [0, \infty]$ are measures on $\mathcal{M}$ for $n \in \mathbb{N}$. Also suppose that $\mu_n(A)$ is increasing in n for all $A \in \mathcal{M}$. Prove that $\mu : \mathcal{M} \to [0, \infty]$ defined by $\mu(A) := \lim_{n\to\infty} \mu_n(A)$ is also a measure.

Exercise 7.20. Now suppose that Λ is some index set and for each $\lambda \in \Lambda$, $\mu_\lambda : \mathcal{M} \to [0, \infty]$ is a measure on $\mathcal{M}$. Define $\mu : \mathcal{M} \to [0, \infty]$ by $\mu(A) = \sum_{\lambda \in \Lambda} \mu_\lambda(A)$ for each $A \in \mathcal{M}$. Show that μ is also a measure.

Exercise 7.21. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show

$$\mu(\{A_n \text{ a.a.}\}) \leq \liminf_{n\to\infty} \mu(A_n)$$

and if $\mu(\cup_{m \geq n} A_m) < \infty$ for some n, then

$$\mu(\{A_n \text{ i.o.}\}) \geq \limsup_{n\to\infty} \mu(A_n).$$

Exercise 7.22 (Folland 2.13 on p. 52.). Suppose that $\{f_n\}_{n=1}^\infty$ is a sequence of non-negative measurable functions such that $f_n \to f$ pointwise and

$$\lim_{n\to\infty} \int f_n = \int f < \infty.$$

Then

$$\int_E f = \lim_{n\to\infty} \int_E f_n$$

for all measurable sets $E \in \mathcal{M}$. The conclusion need not hold if $\lim_{n\to\infty} \int f_n = \int f$. **Hint:** "Fatou times two."

Exercise 7.23. Give examples of measurable functions $\{f_n\}$ on $\mathbb{R}$ such that f_n decreases to 0 uniformly yet $\int f_n dm = \infty$ for all n. Also give an example of a sequence of measurable functions $\{g_n\}$ on $[0, 1]$ such that $g_n \to 0$ while $\int g_n dm = 1$ for all n.

Exercise 7.24. Suppose $\{a_n\}_{n=-\infty}^{\infty} \subset \mathbb{C}$ is a summable sequence (i.e. $\sum_{n=-\infty}^{\infty} |a_n| < \infty$), then $f(\theta) := \sum_{n=-\infty}^{\infty} a_n e^{in\theta}$ is a continuous function for $\theta \in \mathbb{R}$ and

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$

Exercise 7.25. For any function $f \in L^1(m)$, show $x \in \mathbb{R} \to \int_{(-\infty,x]} f(t)\, dm(t)$ is continuous in x. Also find a finite measure, μ, on $\mathcal{B}_{\mathbb{R}}$ such that $x \to \int_{(-\infty,x]} f(t)\, d\mu(t)$ is not continuous.

Exercise 7.26. Folland 2.31b and 2.31e on p. 60. (The answer in 2.13b is wrong by a factor of -1 and the sum is on $k = 1$ to ∞. In part (e), s should be taken to be a. You may also freely use the Taylor series expansion

$$(1-z)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{2^n n!} z^n = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} z^n \text{ for } |z| < 1.$$

Exercise 7.27. Prove Lemma 7.58.

Functional Forms of the $\pi - \lambda$ Theorem

In this chapter we will develop a very useful function analogue of the $\pi - \lambda$ theorem. The results in this section will be used often in the sequel.

Notation 8.1 *Let Ω be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on Ω. We say that $\mathbb{H}$ is **closed under bounded convergence** if; for every sequence, $\{f_n\}_{n=1}^{\infty} \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $|f_n(\omega)| \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f(\omega) := \lim_{n\to\infty} f_n(\omega)$ exists for all $\omega \in \Omega$, then $f \in \mathbb{H}$.*

*A subset, $\mathbb{M}$, of $\mathbb{H}$ is called a **multiplicative system** if $\mathbb{M}$ is closed under finite intersections.*

The following result may be found in Dellacherie [7, p. 14]. The style of proof given here may be found in Janson [20, Appendix A., p. 309].

Theorem 8.2 (Dynkin's Multiplicative System Theorem). *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from Ω to $\mathbb{R}$ which contains the constant functions and is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. First Proof. In this proof, we may (and do) assume that $\mathbb{H}$ is the smallest subspace of bounded functions on Ω which contains the constant functions, contains $\mathbb{M}$, and is closed under bounded convergence. (As usual such a space exists by taking the intersection of all such spaces.) The remainder of the proof will be broken into four steps.

Step 1. ($\mathbb{H}$ is an algebra of functions.) For $f \in \mathbb{H}$, let $\mathbb{H}^f := \{g \in \mathbb{H} : gf \in \mathbb{H}\}$. The reader will now easily verify that $\mathbb{H}^f$ is a linear subspace of $\mathbb{H}$, $1 \in \mathbb{H}^f$, and $\mathbb{H}^f$ is closed under bounded convergence. Moreover if $f \in \mathbb{M}$, since $\mathbb{M}$ is a multiplicative system, $\mathbb{M} \subset \mathbb{H}^f$. Hence by the definition of $\mathbb{H}$, $\mathbb{H} = \mathbb{H}^f$, i.e. $fg \in \mathbb{H}$ for all $f \in \mathbb{M}$ and $g \in \mathbb{H}$. Having proved this it now follows for any $f \in \mathbb{H}$ that $\mathbb{M} \subset \mathbb{H}^f$ and therefore as before, $\mathbb{H}^f = \mathbb{H}$. Thus we may conclude that $fg \in \mathbb{H}$ whenever $f, g \in \mathbb{H}$, i.e. $\mathbb{H}$ is an algebra of functions.

Step 2. ($\mathcal{B} := \{A \subset \Omega : 1_A \in \mathbb{H}\}$ is a σ – algebra.) Using the fact that $\mathbb{H}$ is an algebra containing constants, the reader will easily verify that $\mathcal{B}$ is closed under complementation, finite intersections, and contains Ω, i.e. $\mathcal{B}$ is an algebra. Using the fact that $\mathbb{H}$ is closed under bounded convergence, it follows that $\mathcal{B}$ is closed under increasing unions and hence that $\mathcal{B}$ is σ – algebra.

Step 3. ($\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.) Since $\mathbb{H}$ is a vector space and $\mathbb{H}$ contains 1_A for all $A \in \mathcal{B}$, $\mathbb{H}$ contains all $\mathcal{B}$ – measurable simple functions. Since every bounded $\mathcal{B}$ – measurable function may be written as a bounded limit of such simple functions (see Theorem 6.39), it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions.

Step 4. ($\sigma(\mathbb{M}) \subset \mathcal{B}$.) Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 8.1 below) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f \in \mathbb{M}$ and $a \in \mathbb{R}$, let $F_n := \varphi_n(f - a)$ and $M := \sup_{\omega\in\Omega} |f(\omega) - a|$. By the Weierstrass approximation Theorem 4.36, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$. Since p_l is a polynomial and $\mathbb{H}$ is an algebra, $p_l(f - a) \in \mathbb{H}$ for all l. Moreover, $p_l \circ (f - a) \to F_n$ uniformly as $l \to \infty$, from with it follows that $F_n \in \mathbb{H}$ for all n. Since, $F_n \uparrow 1_{\{f>a\}}$ it follows that $1_{\{f>a\}} \in \mathbb{H}$, i.e. $\{f > a\} \in \mathcal{B}$. As the sets $\{f > a\}$ with $a \in \mathbb{R}$ and $f \in \mathbb{M}$ generate $\sigma(\mathbb{M})$, it follows that $\sigma(\mathbb{M}) \subset \mathcal{B}$.

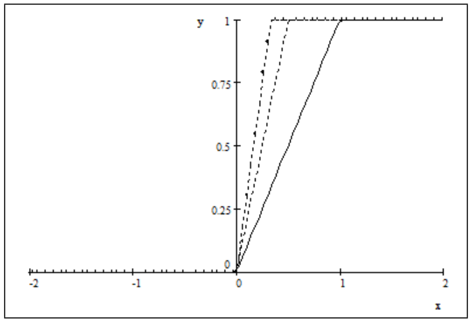


Fig. 8.1. Plots of φ_1, φ_2 and φ_3.

Second proof. (This proof may safely be skipped.) This proof will make use of Dynkin's $\pi - \lambda$ Theorem 5.14. Let

$$\mathcal{L} := \{A \subset \Omega : 1_A \in \mathbb{H}\}.$$

We then have $\Omega \in \mathcal{L}$ since $1_\Omega = 1 \in \mathbb{H}$, if $A, B \in \mathcal{L}$ with $A \subset B$ then $B \setminus A \in \mathcal{L}$ since $1_{B\setminus A} = 1_B - 1_A \in \mathbb{H}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $A \in \mathcal{L}$ because $1_{A_n} \in \mathbb{H}$ and $1_{A_n} \uparrow 1_A \in \mathbb{H}$. Therefore $\mathcal{L}$ is λ – system.

Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 8.1 above) so that $\varphi_n(x) \uparrow 1_{x>0}$. Given $f_1, f_2, \dots, f_k \in \mathbb{M}$ and $a_1, \dots, a_k \in \mathbb{R}$, let

$$F_n := \prod_{i=1}^{k} \varphi_n(f_i - a_i)$$

and let

$$M := \sup_{i=1,\dots,k} \sup_{\omega} |f_i(\omega) - a_i| .$$

By the Weierstrass approximation Theorem 4.36, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$ .Since p_l is a polynomial it is easily seen that $\prod_{i=1}^{k} p_l \circ (f_i - a_i) \in \mathbb{H}$. Moreover,

$$\prod_{i=1}^{k} p_l \circ (f_i - a_i) \to F_n \text{ uniformly as } l \to \infty,$$

from with it follows that $F_n \in \mathbb{H}$ for all n. Since,

$$F_n \uparrow \prod_{i=1}^{k} 1_{\{f_i > a_i\}} = 1_{\cap_{i=1}^{k}\{f_i > a_i\}}$$

it follows that $1_{\cap_{i=1}^{k}\{f_i > a_i\}} \in \mathbb{H}$ or equivalently that $\cap_{i=1}^{k}\{f_i > a_i\} \in \mathcal{L}$. Therefore $\mathcal{L}$ contains the π – system, $\mathcal{P}$, consisting of finite intersections of sets of the form, $\{f > a\}$ with $f \in \mathbb{M}$ and $a \in \mathbb{R}$.

As a consequence of the above paragraphs and the $\pi - \lambda$ Theorem 5.14, $\mathcal{L}$ contains $\sigma(\mathcal{P}) = \sigma(\mathbb{M})$. In particular it follows that $1_A \in \mathbb{H}$ for all $A \in \sigma(\mathbb{M})$. Since any positive $\sigma(\mathbb{M})$ – measurable function may be written as a increasing limit of simple functions (see Theorem 6.39)), it follows that $\mathbb{H}$ contains all non-negative bounded $\sigma(\mathbb{M})$ – measurable functions. Finally, since any bounded $\sigma(\mathbb{M})$ – measurable functions may be written as the difference of two such non-negative simple functions, it follows that $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions. ∎

Here is a complex version of the previous theorem.

Theorem 8.3 (Complex Multiplicative System Theorem). *Suppose $\mathbb{H}$ is a complex linear subspace of the bounded complex functions on Ω, $1 \in \mathbb{H}$, $\mathbb{H}$ is closed under complex conjugation, and $\mathbb{H}$ is closed under bounded convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system which is closed under conjugation, then $\mathbb{H}$ contains all bounded complex valued $\sigma(\mathbb{M})$-measurable functions.*

Proof. Let $\mathbb{M}_0 = \mathrm{span}_{\mathbb{C}}(\mathbb{M} \cup \{1\})$ be the complex span of $\mathbb{M}$. As the reader should verify, $\mathbb{M}_0$ is an algebra, $\mathbb{M}_0 \subset \mathbb{H}$, $\mathbb{M}_0$ is closed under complex conjugation and $\sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$. Let

$$\mathbb{H}^{\mathbb{R}} := \{f \in \mathbb{H} : f \text{ is real valued}\} \text{ and}$$
$$\mathbb{M}_0^{\mathbb{R}} := \{f \in \mathbb{M}_0 : f \text{ is real valued}\} .$$

Then $\mathbb{H}^{\mathbb{R}}$ is a real linear space of bounded real valued functions 1 which is closed under bounded convergence and $\mathbb{M}_0^{\mathbb{R}} \subset \mathbb{H}^{\mathbb{R}}$. Moreover, $\mathbb{M}_0^{\mathbb{R}}$ is a multiplicative system (as the reader should check) and therefore by Theorem 8.6, $\mathbb{H}^{\mathbb{R}}$ contains all bounded $\sigma(\mathbb{M}_0^{\mathbb{R}})$ – measurable real valued functions. Since $\mathbb{H}$ and $\mathbb{M}_0$ are complex linear spaces closed under complex conjugation, for any $f \in \mathbb{H}$ or $f \in \mathbb{M}_0$, the functions $\mathrm{Re}\, f = \frac{1}{2}(f + \bar{f})$ and $\mathrm{Im}\, f = \frac{1}{2i}(f - \bar{f})$ are in $\mathbb{H}$ or $\mathbb{M}_0$ respectively. Therefore $\mathbb{M}_0 = \mathbb{M}_0^{\mathbb{R}} + i\mathbb{M}_0^{\mathbb{R}}$, $\sigma(\mathbb{M}_0^{\mathbb{R}}) = \sigma(\mathbb{M}_0) = \sigma(\mathbb{M})$, and $\mathbb{H} = \mathbb{H}^{\mathbb{R}} + i\mathbb{H}^{\mathbb{R}}$. Hence if $f : \Omega \to \mathbb{C}$ is a bounded $\sigma(\mathbb{M})$ – measurable function, then $f = \mathrm{Re}\, f + i\, \mathrm{Im}\, f \in \mathbb{H}$ since $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ are in $\mathbb{H}^{\mathbb{R}}$. ∎

Notation 8.4 *We say that $\mathbb{H} \subset \ell^\infty(\Omega, \mathbb{R})$ is* ***closed under monotone convergence*** *if; for every sequence, $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$, satisfying:*

1. *there exists $M < \infty$ such that $0 \le f_n(\omega) \le M$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$,*
2. *$f_n(\omega)$ is increasing in n for all $\omega \in \Omega$, then $f := \lim_{n\to\infty} f_n \in \mathbb{H}$.*

Clearly if $\mathbb{H}$ is closed under bounded convergence then it is also closed under monotone convergence. I learned the proof of the converse from Pat Fitzsimmons but this result appears in Sharpe [48, p. 365].

Proposition 8.5. *Let Ω be a set. Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions from Ω to $\mathbb{R}$ which is closed under monotone convergence. Then $\mathbb{H}$ is closed under uniform convergence as well, i.e. $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ with $\sup_{n\in\mathbb{N}} \sup_{\omega\in\Omega} |f_n(\omega)| < \infty$ and $f_n \to f$, then $f \in \mathbb{H}$.*

Proof. Let us first assume that $\{f_n\}_{n=1}^\infty \subset \mathbb{H}$ such that f_n converges uniformly to a bounded function, $f : \Omega \to \mathbb{R}$. Let $\|f\|_\infty := \sup_{\omega\in\Omega} |f(\omega)|$. Let $\varepsilon > 0$ be given. By passing to a subsequence if necessary, we may assume $\|f - f_n\|_\infty \le \varepsilon 2^{-(n+1)}$. Let

$$g_n := f_n - \delta_n + M$$

with δ_n and M constants to be determined shortly. We then have

$$g_{n+1} - g_n = f_{n+1} - f_n + \delta_n - \delta_{n+1} \ge -\varepsilon 2^{-(n+1)} + \delta_n - \delta_{n+1}.$$

Taking $\delta_n := \varepsilon 2^{-n}$, then $\delta_n - \delta_{n+1} = \varepsilon 2^{-n}(1 - 1/2) = \varepsilon 2^{-(n+1)}$ in which case $g_{n+1} - g_n \ge 0$ for all n. By choosing M sufficiently large, we will also have $g_n \ge 0$ for all n. Since $\mathbb{H}$ is a vector space containing the constant functions, $g_n \in \mathbb{H}$ and since $g_n \uparrow f + M$, it follows that $f = f + M - M \in \mathbb{H}$. So we have shown that $\mathbb{H}$ is closed under uniform convergence. ∎

This proposition immediately leads to the following strengthening of Theorem 8.2.

Theorem 8.6. *Suppose that $\mathbb{H}$ is a vector subspace of bounded functions from Ω to $\mathbb{R}$ which contains the constant functions and is closed under monotone convergence. If $\mathbb{M} \subset \mathbb{H}$ is multiplicative system, then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. Proposition 8.5 reduces this theorem to Theorem 8.2. ∎

Exercise 8.1. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be a pair of random variables such that

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)g(X)]$$

for every pair of bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$. Show $P(X = Y) = 1$. **Hint:** Let $\mathbb{H}$ denote the bounded Borel measurable functions, $h : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbb{E}[h(X,Y)] = \mathbb{E}[h(X,X)].$$

Use Theorem 8.2 to show $\mathbb{H}$ is the vector space of all bounded Borel measurable functions. Then take $h(x,y) = 1_{\{x=y\}}$.

Corollary 8.7. *Suppose $\mathbb{H}$ is a real subspace of bounded functions such that $1 \in \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. If $\mathcal{P} \subset 2^{\Omega}$ is a multiplicative class such that $1_A \in \mathbb{H}$ for all $A \in \mathcal{P}$, then $\mathbb{H}$ contains all bounded $\sigma(\mathcal{P})$ – measurable functions.*

Proof. Let $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$. Then $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system and the proof is completed with an application of Theorem 8.6. ∎

Example 8.8. Suppose μ and ν are two probability measure on $(\Omega, \mathcal{B})$ such that

$$\int_\Omega f d\mu = \int_\Omega f d\nu \tag{8.1}$$

for all f in a multiplicative subset, $\mathbb{M}$, of bounded measurable functions on Ω. Then $\mu = \nu$ on $\sigma(\mathbb{M})$. Indeed, apply Theorem 8.6 with $\mathbb{H}$ being the bounded measurable functions on Ω such that Eq. (8.1) holds. In particular if $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$ with $\mathcal{P}$ being a multiplicative class we learn that $\mu = \nu$ on $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

Corollary 8.9. *The smallest subspace of real valued functions, $\mathbb{H}$, on $\mathbb{R}$ which contains $C_c(\mathbb{R}, \mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ with compact support) is the collection of bounded Borel measurable function on $\mathbb{R}$.*

Proof. By a homework problem, for $-\infty < a < b < \infty$, $1_{(a,b]}$ may be written as a bounded limit of continuous functions with compact support from which it follows that $\sigma(C_c(\mathbb{R}, \mathbb{R})) = \mathcal{B}_{\mathbb{R}}$. It is also easy to see that 1 is a bounded limit of functions in $C_c(\mathbb{R}, \mathbb{R})$ and hence $1 \in \mathbb{H}$. The corollary now follows by an application of The result now follows by an application of Theorem 8.6 with $\mathbb{M} := C_c(\mathbb{R}, \mathbb{R})$. ∎

8.0.1 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

Notation 8.10 *Given a collection of bounded functions, $\mathbb{M}$, from a set, Ω, to $\mathbb{R}$, let $\mathbb{M}_\uparrow$ ($\mathbb{M}_\downarrow$) denote the the bounded monotone increasing (decreasing) limits of functions from $\mathbb{M}$. More explicitly a bounded function, $f : \Omega \to \mathbb{R}$ is in $\mathbb{M}_\uparrow$ respectively $\mathbb{M}_\downarrow$ iff there exists $f_n \in \mathbb{M}$ such that $f_n \uparrow f$ respectively $f_n \downarrow f$.*

Theorem 8.11 (Bounded Approximation Theorem*). *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that:*

1. *$\sigma(\mathbb{M}) = \mathcal{B}$,*
2. *$1 \in \mathbb{M}$, and*
3. *$|f| \in \mathbb{M}$ for all $f \in \mathbb{M}$.*

Then for every bounded $\sigma(\mathbb{M})$ measurable function, $g : \Omega \to \mathbb{R}$, and every $\varepsilon > 0$, there exists $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$.[1]

Proof. Let us begin with a few simple observations.

1. $\mathbb{M}$ is a "lattice" – if $f, g \in \mathbb{M}$ then
$$f \vee g = \frac{1}{2}(f + g + |f - g|) \in \mathbb{M}$$
and
$$f \wedge g = \frac{1}{2}(f + g - |f - g|) \in \mathbb{M}.$$
2. If $f, g \in \mathbb{M}_\uparrow$ or $f, g \in \mathbb{M}_\downarrow$ then $f + g \in \mathbb{M}_\uparrow$ or $f + g \in \mathbb{M}_\downarrow$ respectively.
3. If $\lambda \ge 0$ and $f \in \mathbb{M}_\uparrow$ ($f \in \mathbb{M}_\downarrow$), then $\lambda f \in \mathbb{M}_\uparrow$ ($\lambda f \in \mathbb{M}_\downarrow$).
4. If $f \in \mathbb{M}_\uparrow$ then $-f \in \mathbb{M}_\downarrow$ and visa versa.
5. If $f_n \in \mathbb{M}_\uparrow$ and $f_n \uparrow f$ where $f : \Omega \to \mathbb{R}$ is a bounded function, then $f \in \mathbb{M}_\uparrow$. Indeed, by assumption there exists $f_{n,i} \in \mathbb{M}$ such that $f_{n,i} \uparrow f_n$ as $i \to \infty$. By observation (1), $g_n := \max\{f_{ij} : i, j \le n\} \in \mathbb{M}$. Moreover it is clear that $g_n \le \max\{f_k : k \le n\} = f_n \le f$ and hence $g_n \uparrow g := \lim_{n\to\infty} g_n \le f$. Since $f_{ij} \le g$ for all i, j, it follows that $f_n = \lim_{j\to\infty} f_{nj} \le g$ and consequently that $f = \lim_{n\to\infty} f_n \le g \le f$. So we have shown that $g_n \uparrow f \in \mathbb{M}_\uparrow$.

Now let $\mathbb{H}$ denote the collection of bounded measurable functions which satisfy the assertion of the theorem. Clearly, $\mathbb{M} \subset \mathbb{H}$ and in fact it is also easy to see that $\mathbb{M}_\uparrow$ and $\mathbb{M}_\downarrow$ are contained in $\mathbb{H}$ as well. For example, if $f \in \mathbb{M}_\uparrow$, by definition, there exists $f_n \in \mathbb{M} \subset \mathbb{M}_\downarrow$ such that $f_n \uparrow f$. Since $\mathbb{M}_\downarrow \ni f_n \le f \le$

[1] Bruce: rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

$f \in \mathbb{M}_\uparrow$ and $\mu(f - f_n) \to 0$ by the dominated convergence theorem, it follows that $f \in \mathbb{H}$. As similar argument shows $\mathbb{M}_\downarrow \subset \mathbb{H}$. We will now show $\mathbb{H}$ is a vector sub-space of the bounded $\mathcal{B} = \sigma(\mathbb{M})$ – measurable functions.

$\mathbb{H}$ **is closed under addition.** If $g_i \in \mathbb{H}$ for $i = 1, 2$, and $\varepsilon > 0$ is given, we may find $f_i \in \mathbb{M}_\downarrow$ and $h_i \in \mathbb{M}_\uparrow$ such that $f_i \le g_i \le h_i$ and $\mu(h_i - f_i) < \varepsilon/2$ for $i = 1, 2$. Since $h = h_1 + h_2 \in \mathbb{M}_\uparrow$, $f := f_1 + f_2 \in \mathbb{M}_\downarrow$, $f \le g_1 + g_2 \le h$, and

$$\mu(h - f) = \mu(h_1 - f_1) + \mu(h_2 - f_2) < \varepsilon,$$

it follows that $g_1 + g_2 \in \mathbb{H}$.

$\mathbb{H}$ **is closed under scalar multiplication.** If $g \in \mathbb{H}$ then $\lambda g \in \mathbb{H}$ for all $\lambda \in \mathbb{R}$. Indeed suppose that $\varepsilon > 0$ is given and $f \in \mathbb{M}_\downarrow$ and $h \in \mathbb{M}_\uparrow$ such that $f \le g \le h$ and $\mu(h - f) < \varepsilon$. Then for $\lambda \ge 0$, $\mathbb{M}_\downarrow \ni \lambda f \le \lambda g \le \lambda h \in \mathbb{M}_\uparrow$ and

$$\mu(\lambda h - \lambda f) = \lambda \mu(h - f) < \lambda \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, if follows that $\lambda g \in \mathbb{H}$ for $\lambda \ge 0$. Similarly, $\mathbb{M}_\downarrow \ni -h \le -g \le -f \in \mathbb{M}_\uparrow$ and

$$\mu(-f - (-h)) = \mu(h - f) < \varepsilon.$$

which shows $-g \in \mathbb{H}$ as well.

Because of Theorem 8.6, to complete this proof, it suffices to show $\mathbb{H}$ is closed under monotone convergence. So suppose that $g_n \in \mathbb{H}$ and $g_n \uparrow g$, where $g : \Omega \to \mathbb{R}$ is a bounded function. Since $\mathbb{H}$ is a vector space, it follows that $0 \le \delta_n := g_{n+1} - g_n \in \mathbb{H}$ for all $n \in \mathbb{N}$. So if $\varepsilon > 0$ is given, we can find, $\mathbb{M}_\downarrow \ni u_n \le \delta_n \le v_n \in \mathbb{M}_\uparrow$ such that $\mu(v_n - u_n) \le 2^{-n}\varepsilon$ for all n. By replacing u_n by $u_n \vee 0 \in \mathbb{M}_\downarrow$ (by observation 1.), we may further assume that $u_n \ge 0$. Let

$$v := \sum_{n=1}^\infty v_n =\uparrow \lim_{N\to\infty} \sum_{n=1}^N v_n \in \mathbb{M}_\uparrow \text{ (using observations 2. and 5.)}$$

and for $N \in \mathbb{N}$, let

$$u^N := \sum_{n=1}^N u_n \in \mathbb{M}_\downarrow \text{ (using observation 2).}$$

Then

$$\sum_{n=1}^\infty \delta_n = \lim_{N\to\infty} \sum_{n=1}^N \delta_n = \lim_{N\to\infty} (g_{N+1} - g_1) = g - g_1$$

and $u^N \le g - g_1 \le v$. Moreover,

$$\mu\left(v - u^N\right) = \sum_{n=1}^N \mu(v_n - u_n) + \sum_{n=N+1}^\infty \mu(v_n) \le \sum_{n=1}^N \varepsilon 2^{-n} + \sum_{n=N+1}^\infty \mu(v_n)$$
$$\le \varepsilon + \sum_{n=N+1}^\infty \mu(v_n).$$

However, since

$$\sum_{n=1}^\infty \mu(v_n) \le \sum_{n=1}^\infty \mu\left(\delta_n + \varepsilon 2^{-n}\right) = \sum_{n=1}^\infty \mu(\delta_n) + \varepsilon\mu(\Omega)$$
$$= \sum_{n=1}^\infty \mu(g - g_1) + \varepsilon\mu(\Omega) < \infty,$$

it follows that for $N \in \mathbb{N}$ sufficiently large that $\sum_{n=N+1}^\infty \mu(v_n) < \varepsilon$. Therefore, for this N, we have $\mu\left(v - u^N\right) < 2\varepsilon$ and since $\varepsilon > 0$ is arbitrary, if follows that $g - g_1 \in \mathbb{H}$. Since $g_1 \in \mathbb{H}$ and $\mathbb{H}$ is a vector space, we may conclude that $g = (g - g_1) + g_1 \in \mathbb{H}$. ■

9

Multiple and Iterated Integrals

9.1 Iterated Integrals

Notation 9.1 (Iterated Integrals) *If $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are two measure spaces and $f: X\times Y\to\mathbb{C}$ is a $\mathcal{M}\otimes\mathcal{N}$ – measurable function, the **iterated integrals** of f (when they make sense) are:*

$$\int_X d\mu(x)\int_Y d\nu(y)f(x,y) := \int_X\left[\int_Y f(x,y)d\nu(y)\right]d\mu(x)$$

and

$$\int_Y d\nu(y)\int_X d\mu(x)f(x,y) := \int_Y\left[\int_X f(x,y)d\mu(x)\right]d\nu(y).$$

Notation 9.2 *Suppose that $f: X\to\mathbb{C}$ and $g: Y\to\mathbb{C}$ are functions, let $f\otimes g$ denote the function on $X\times Y$ given by*

$$f\otimes g(x,y) = f(x)g(y).$$

Notice that if f,g are measurable, then $f\otimes g$ is $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{C}})$ – measurable. To prove this let $F(x,y)=f(x)$ and $G(x,y)=g(y)$ so that $f\otimes g = F\cdot G$ will be measurable provided that F and G are measurable. Now $F = f\circ\pi_1$ where $\pi_1: X\times Y\to X$ is the projection map. This shows that F is the composition of measurable functions and hence measurable. Similarly one shows that G is measurable.

9.2 Tonelli's Theorem and Product Measure

Theorem 9.3. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ-finite measure spaces and f is a nonnegative $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable function, then for each $y\in Y$,*

$$x\to f(x,y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.1}$$

for each $x\in X$,

$$y\to f(x,y) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.2}$$

$$x\to\int_Y f(x,y)d\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.3}$$

$$y\to\int_X f(x,y)d\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{9.4}$$

and

$$\int_X d\mu(x)\int_Y d\nu(y)f(x,y) = \int_Y d\nu(y)\int_X d\mu(x)f(x,y). \tag{9.5}$$

Proof. Suppose that $E = A\times B\in\mathcal{E} := \mathcal{M}\times\mathcal{N}$ and $f = 1_E$. Then

$$f(x,y) = 1_{A\times B}(x,y) = 1_A(x)1_B(y)$$

and one sees that Eqs. (9.1) and (9.2) hold. Moreover

$$\int_Y f(x,y)d\nu(y) = \int_Y 1_A(x)1_B(y)d\nu(y) = 1_A(x)\nu(B),$$

so that Eq. (9.3) holds and we have

$$\int_X d\mu(x)\int_Y d\nu(y)f(x,y) = \nu(B)\mu(A). \tag{9.6}$$

Similarly,

$$\int_X f(x,y)d\mu(x) = \mu(A)1_B(y) \text{ and}$$
$$\int_Y d\nu(y)\int_X d\mu(x)f(x,y) = \nu(B)\mu(A)$$

from which it follows that Eqs. (9.4) and (9.5) hold in this case as well.

For the moment let us now further assume that $\mu(X)<\infty$ and $\nu(Y)<\infty$ and let $\mathbb{H}$ be the collection of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$ such that Eqs. (9.1) – (9.5) hold. Using the fact that measurable functions are closed under pointwise limits and the dominated convergence theorem (the dominating function always being a constant), one easily shows that $\mathbb{H}$ closed under bounded convergence. Since we have just verified that $1_E\in\mathbb{H}$ for all E in the π – class, $\mathcal{E}$, it follows by Corollary 8.7 that $\mathbb{H}$ is the space of all bounded $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\mathbb{R}})$ – measurable functions on $X\times Y$. Moreover, if $f:X\times Y\to[0,\infty]$ is a $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, let $f_M = M\wedge f$ so that $f_M\uparrow f$ as $M\to\infty$. Then Eqs. (9.1) – (9.5) hold with f replaced by f_M for all $M\in\mathbb{N}$. Repeated use of the monotone convergence theorem allows us to pass to the limit $M\to\infty$ in these equations to deduce the theorem in the case μ and ν are finite measures.

For the σ – finite case, choose $X_n\in\mathcal{M}$, $Y_n\in\mathcal{N}$ such that $X_n\uparrow X$, $Y_n\uparrow Y$, $\mu(X_n)<\infty$ and $\nu(Y_n)<\infty$ for all $m,n\in\mathbb{N}$. Then define $\mu_m(A)=\mu(X_m\cap A)$ and $\nu_n(B)=\nu(Y_n\cap B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ or equivalently $d\mu_m = 1_{X_m}d\mu$ and $d\nu_n = 1_{Y_n}d\nu$. By what we have just proved Eqs. (9.1) – (9.5) with μ replaced by μ_m and ν by ν_n for all $(\mathcal{M}\otimes\mathcal{N},\mathcal{B}_{\bar{\mathbb{R}}})$ – measurable functions, $f:X\times Y\to[0,\infty]$. The validity of Eqs. (9.1) – (9.5) then follows by passing to the limits $m\to\infty$ and then $n\to\infty$ making use of the monotone convergence theorem in the following context. For all $u\in L^+(X,\mathcal{M})$,

$$\int_X u\,d\mu_m = \int_X u 1_{X_m}d\mu \uparrow \int_X u\,d\mu \text{ as } m\to\infty,$$

and for all and $v\in L^+(Y,\mathcal{N})$,

$$\int_Y v\,d\mu_n = \int_Y v 1_{Y_n}d\mu \uparrow \int_Y v\,d\mu \text{ as } n\to\infty.$$

■

Corollary 9.4. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces. Then there exists a unique measure π on $\mathcal{M}\otimes\mathcal{N}$ such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$. Moreover π is given by*

$$\pi(E)=\int_X d\mu(x)\int_Y d\nu(y)1_E(x,y)=\int_Y d\nu(y)\int_X d\mu(x)1_E(x,y) \tag{9.7}$$

for all $E\in\mathcal{M}\otimes\mathcal{N}$ and π is σ – finite.

Proof. Notice that any measure π such that $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ is necessarily σ – finite. Indeed, let $X_n\in\mathcal{M}$ and $Y_n\in\mathcal{N}$ be chosen so that $\mu(X_n)<\infty$, $\nu(Y_n)<\infty$, $X_n\uparrow X$ and $Y_n\uparrow Y$, then $X_n\times Y_n\in\mathcal{M}\otimes\mathcal{N}$, $X_n\times Y_n\uparrow X\times Y$ and $\pi(X_n\times Y_n)<\infty$ for all n. The uniqueness assertion is a consequence of the combination of Exercises 3.10 and 5.11 Proposition 3.25 with $\mathcal{E}=\mathcal{M}\times\mathcal{N}$. For the existence, it suffices to observe, using the monotone convergence theorem, that π defined in Eq. (9.7) is a measure on $\mathcal{M}\otimes\mathcal{N}$. Moreover this measure satisfies $\pi(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{M}$ and $B\in\mathcal{N}$ from Eq. (9.6). ■

Notation 9.5 *The measure π is called the product measure of μ and ν and will be denoted by $\mu\otimes\nu$.*

Theorem 9.6 (Tonelli's Theorem). *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are σ – finite measure spaces and $\pi=\mu\otimes\nu$ is the product measure on $\mathcal{M}\otimes\mathcal{N}$. If $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$, then $f(\cdot,y)\in L^+(X,\mathcal{M})$ for all $y\in Y$, $f(x,\cdot)\in L^+(Y,\mathcal{N})$ for all $x\in X$,*

$$\int_Y f(\cdot,y)d\nu(y)\in L^+(X,\mathcal{M}),\ \int_X f(x,\cdot)d\mu(x)\in L^+(Y,\mathcal{N})$$

and

$$\int_{X\times Y} f\,d\pi = \int_X d\mu(x)\int_Y d\nu(y)f(x,y) \tag{9.8}$$
$$= \int_Y d\nu(y)\int_X d\mu(x)f(x,y). \tag{9.9}$$

Proof. By Theorem 9.3 and Corollary 9.4, the theorem holds when $f=1_E$ with $E\in\mathcal{M}\otimes\mathcal{N}$. Using the linearity of all of the statements, the theorem is also true for non-negative simple functions. Then using the monotone convergence theorem repeatedly along with the approximation Theorem 6.39, one deduces the theorem for general $f\in L^+(X\times Y,\mathcal{M}\otimes\mathcal{N})$. ■

Example 9.7. In this example we are going to show, $I:=\int_{\mathbb{R}}e^{-x^2/2}dm(x)=\sqrt{2\pi}$. To this end we observe, using Tonelli's theorem, that

$$I^2=\left[\int_{\mathbb{R}}e^{-x^2/2}dm(x)\right]^2=\int_{\mathbb{R}}e^{-y^2/2}\left[\int_{\mathbb{R}}e^{-x^2/2}dm(x)\right]dm(y)$$
$$=\int_{\mathbb{R}^2}e^{-(x^2+y^2)/2}dm^2(x,y)$$

where $m^2=m\otimes m$ is "Lebesgue measure" on $\left(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2}=\mathcal{B}_{\mathbb{R}}\otimes\mathcal{B}_{\mathbb{R}}\right)$. From the monotone convergence theorem,

$$I^2=\lim_{R\to\infty}\int_{D_R}e^{-(x^2+y^2)/2}d\pi(x,y)$$

where $D_R=\left\{(x,y):x^2+y^2<R^2\right\}$. Using the change of variables theorem described in Section 9.5 below,[1] we find

$$\int_{D_R}e^{-(x^2+y^2)/2}d\pi(x,y)=\int_{(0,R)\times(0,2\pi)}e^{-r^2/2}r\,dr\,d\theta$$
$$=2\pi\int_0^R e^{-r^2/2}r\,dr=2\pi\left(1-e^{-R^2/2}\right).$$

[1] Alternatively, you can easily show that the integral $\int_{D_R}f\,dm^2$ agrees with the multiple integral in undergraduate analysis when f is continuous. Then use the change of variables theorem from undergraduate analysis.

From this we learn that

$$I^2 = \lim_{R\to\infty} 2\pi \left(1 - e^{-R^2/2}\right) = 2\pi$$

as desired.

9.3 Fubini's Theorem

The following convention will be in force for the rest of this section.

Convention: If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is a measurable but non-integrable function, i.e. $\int_X |f|\, d\mu = \infty$, by convention we will define $\int_X f d\mu := 0$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) is a non-integrable function we will still write $\int_X f d\mu = \infty$.

Theorem 9.8 (Fubini's Theorem). *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ – finite measure spaces, $\pi = \mu \otimes \nu$ is the product measure on $\mathcal{M} \otimes \mathcal{N}$ and $f : X \times Y \to \mathbb{C}$ is a $\mathcal{M} \otimes \mathcal{N}$ – measurable function. Then the following three conditions are equivalent:*

$$\int_{X\times Y} |f|\, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{9.10}$$

$$\int_X \left(\int_Y |f(x, y)|\, d\nu(y)\right) d\mu(x) < \infty \text{ and} \tag{9.11}$$

$$\int_Y \left(\int_X |f(x, y)|\, d\mu(x)\right) d\nu(y) < \infty. \tag{9.12}$$

If any one (and hence all) of these condition hold, then $f(x, \cdot) \in L^1(\nu)$ for μ-a.e. x, $f(\cdot, y) \in L^1(\mu)$ for ν-a.e. y, $\int_Y f(\cdot, y)dv(y) \in L^1(\mu)$, $\int_X f(x, \cdot)d\mu(x) \in L^1(\nu)$ and Eqs. (9.8) and (9.9) are still valid.

Proof. The equivalence of Eqs. (9.10) – (9.12) is a direct consequence of Tonelli's Theorem 9.6. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{x \in X : \int_Y |f(x, y)|\, d\nu(y) = \infty\right\}. \tag{9.13}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x, y)|\, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

$$\int_X \left[\int_Y |f(x, y)|\, d\nu(y)\right] d\mu(x) = \int_{X\times Y} |f|\, d\pi < \infty$$

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of f, then using the above convention we have

$$\begin{aligned}\int_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) [f_+(x, y) - f_-(x, y)]\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x, y)\, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x, y)\, d\nu(y).\end{aligned} \tag{9.14}$$

Noting that $1_{E^c}(x) f_\pm(x, y) = (1_{E^c} \otimes 1_Y \cdot f_\pm)(x, y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \int_Y f(x, y)\, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left|\int_Y f(x, y)\, d\nu(y)\right| d\mu(x) \le \int_X \left[\int_Y |f(x, y)|\, d\nu(y)\right] d\mu(x) < \infty,$$

which shows $\int_Y f(\cdot, y)dv(y) \in L^1(\mu)$. Integrating Eq. (9.14) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned}&\int_X \left[\int_Y f(x, y)\, d\nu(y)\right] d\mu(x) \\ &= \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_+(x, y) - \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) f_-(x, y) \\ &= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-)\, d\pi = \int_{X\times Y} f d\pi\end{aligned} \tag{9.15}$$

which proves Eq. (9.8) holds.

Now suppose that $f = u + iv$ is complex valued and again let E be as in Eq. (9.13). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) = 0$. By our convention,

$$\begin{aligned}\int_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) = \int_Y 1_{E^c}(x) [u(x, y) + iv(x, y)]\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) u(x, y)\, d\nu(y) + i \int_Y 1_{E^c}(x) v(x, y)\, d\nu(y)\end{aligned}$$

which is measurable in x by what we have just proved. Similarly one shows $\int_Y f(\cdot, y)\, d\nu(y) \in L^1(\mu)$ and Eq. (9.8) still holds by a computation similar to that done in Eq. (9.15). The assertions pertaining to Eq. (9.9) may be proved in the same way. ■

The previous theorems have obvious generalizations to products of any finite number of σ – finite measure spaces. For example the following theorem holds.

Theorem 9.9. *Suppose* $\{(X_i, \mathcal{M}_i, \mu_i)\}_{i=1}^n$ *are* σ *– finite measure spaces and* $X := X_1 \times \cdots \times X_n$. *Then there exists a unique measure,* π, *on* $(X, \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n)$ *such that*

$$\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1) \ldots \mu_n(A_n) \text{ for all } A_i \in \mathcal{M}_i.$$

(This measure and its completion will be denoted by $\mu_1 \otimes \cdots \otimes \mu_n$.*) If* $f : X \to [0, \infty]$ *is a* $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ *– measurable function then*

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1, \ldots, x_n) \quad (9.16)$$

where σ *is any permutation of* $\{1, 2, \ldots, n\}$. *This equation also holds for any* $f \in L^1(\pi)$ *and moreover,* $f \in L^1(\pi)$ *iff*

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ |f(x_1, \ldots, x_n)| < \infty$$

for some (and hence all) permutations, σ.

This theorem can be proved by the same methods as in the two factor case, see Exercise 9.4. Alternatively, one can use the theorems already proved and induction on n, see Exercise 9.5 in this regard.

Example 9.10. In this example we will show

$$\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} dx = \pi/2. \quad (9.17)$$

To see this write $\frac{1}{x} = \int_0^\infty e^{-tx} dt$ and use Fubini-Tonelli to conclude that

$$\begin{aligned} \int_0^M \frac{\sin x}{x} dx &= \int_0^M \left[\int_0^\infty e^{-tx} \sin x\ dt\right] dx \\ &= \int_0^\infty \left[\int_0^M e^{-tx} \sin x\ dx\right] dt \\ &= \int_0^\infty \frac{1}{1+t^2} \left(1 - te^{-Mt} \sin M - e^{-Mt} \cos M\right) dt \\ &\to \int_0^\infty \frac{1}{1+t^2} dt = \frac{\pi}{2} \text{ as } M \to \infty, \end{aligned}$$

wherein we have used the dominated convergence theorem (for instance, take $g(t) := \frac{1}{1+t^2}\left(1 + te^{-t} + e^{-t}\right))$ to pass to the limit.

The next example is a refinement of this result.

Example 9.11. We have

$$\int_0^\infty \frac{\sin x}{x} e^{-\Lambda x} dx = \frac{1}{2}\pi - \arctan \Lambda \text{ for all } \Lambda > 0 \quad (9.18)$$

and for$\Lambda, M \in [0, \infty)$,

$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C \frac{e^{-M\Lambda}}{M} \quad (9.19)$$

where $C = \max_{x \ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (9.17) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1 dy\right| = |x|$$

so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x \ne 0$. Making use of the identity

$$\int_0^\infty e^{-tx} dt = 1/x$$

and Fubini's theorem,

$$\begin{aligned} \int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x\, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\ &= \int_0^\infty dt \int_0^M dx \sin x\, e^{-(\Lambda+t)x} \\ &= \int_0^\infty \frac{1 - (\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2 + 1} dt \\ &= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\ &= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M, \Lambda) \end{aligned} \quad (9.20)$$

where

$$\varepsilon(M, \Lambda) = \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt.$$

Since

$$\left|\frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1}\right| \le \frac{1 + (\Lambda+t)}{(\Lambda+t)^2+1} \le C,$$

$$|\varepsilon(M,\Lambda)| \le \int_0^\infty e^{-M(\Lambda+t)}dt = C\frac{e^{-M\Lambda}}{M}.$$

This estimate along with Eq. (9.20) proves Eq. (9.19) from which Eq. (9.17) follows by taking $\Lambda \to \infty$ and Eq. (9.18) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

Lemma 9.12. *Suppose that X is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is a C^1 – functions such that $\lim_{x\to-\infty}\varphi(x) = 0$ and either $\varphi'(x) \ge 0$ for all x or $\int_{\mathbb{R}} |\varphi'(x)| dx < \infty$. Then*

$$\mathbb{E}[\varphi(X)] = \int_{-\infty}^\infty \varphi'(y) P(X > y) dy.$$

Similarly if $X \ge 0$ and $\varphi : [0,\infty) \to \mathbb{R}$ is a C^1 – function such that $\varphi(0) = 0$ and either $\varphi' \ge 0$ or $\int_0^\infty |\varphi'(x)| dx < \infty$, then

$$\mathbb{E}[\varphi(X)] = \int_0^\infty \varphi'(y) P(X > y) dy.$$

Proof. By the fundamental theorem of calculus for all $M < \infty$ and $x \in \mathbb{R}$,

$$\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y) dy. \tag{9.21}$$

Under the stated assumptions on φ, we may use either the monotone or the dominated convergence theorem to let $M \to \infty$ in Eq. (9.21) to find,

$$\varphi(x) = \int_{-\infty}^x \varphi'(y) dy = \int_{\mathbb{R}} 1_{y<x}\varphi'(y) dy \text{ for all } x \in \mathbb{R}.$$

Therefore,

$$\mathbb{E}[\varphi(X)] = \mathbb{E}\left[\int_{\mathbb{R}} 1_{y<X}\varphi'(y) dy\right] = \int_{\mathbb{R}} \mathbb{E}[1_{y<X}]\varphi'(y) dy = \int_{-\infty}^\infty \varphi'(y) P(X > y) dy,$$

where we applied Fubini's theorem for the second equality. The proof of the second assertion is similar and will be left to the reader. ■

Example 9.13. a couple of examples involving Lemma 9.12.

1. Suppose X is a random variable, then
$$\mathbb{E}\left[e^X\right] = \int_{-\infty}^\infty P(X > y) e^y dy = \int_0^\infty P(X > \ln u) du, \tag{9.22}$$
where we made the change of variables, $u = e^y$, to get the second equality.
2. If $X \ge 0$ and $p \ge 1$, then
$$\mathbb{E}X^p = p\int_0^\infty y^{p-1} P(X > y) dy. \tag{9.23}$$

9.4 Fubini's Theorem and Completions*

Notation 9.14 *Given $E \subset X \times Y$ and $x \in X$, let*

$${}_xE := \{y \in Y : (x,y) \in E\}.$$

Similarly if $y \in Y$ is given let

$$E_y := \{x \in X : (x,y) \in E\}.$$

If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.

Theorem 9.15. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are **complete** σ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If f is $\mathcal{L}$ – measurable and (a) $f \ge 0$ or (b) $f \in L^1(\lambda)$ then f_x is $\mathcal{N}$ – measurable for μ a.e. x and f^y is $\mathcal{M}$ – measurable for ν a.e. y and in case (b) $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$ for μ a.e. x and ν a.e. y respectively. Moreover,*

$$\left(x \to \int_Y f_x d\nu\right) \in L^1(\mu) \text{ and } \left(y \to \int_X f^y d\mu\right) \in L^1(\nu)$$

and

$$\int_{X\times Y} f d\lambda = \int_Y d\nu \int_X d\mu\, f = \int_X d\mu \int_Y d\nu\, f.$$

Proof. If $E \in \mathcal{M}\otimes\mathcal{N}$ is a $\mu\otimes\nu$ null set (i.e. $(\mu\otimes\nu)(E) = 0$), then

$$0 = (\mu\otimes\nu)(E) = \int_X \nu({}_xE)d\mu(x) = \int_X \mu(E_y)d\nu(y).$$

This shows that

$$\mu(\{x : \nu({}_xE) \ne 0\}) = 0 \text{ and } \nu(\{y : \mu(E_y) \ne 0\}) = 0,$$

i.e. $\nu({}_xE) = 0$ for μ a.e. x and $\mu(E_y) = 0$ for ν a.e. y. If h is $\mathcal{L}$ measurable and $h = 0$ for λ – a.e., then there exists $E \in \mathcal{M}\otimes\mathcal{N}$ such that $\{(x,y) : h(x,y) \ne 0\} \subset E$ and $(\mu\otimes\nu)(E) = 0$. Therefore $|h(x,y)| \le 1_E(x,y)$ and $(\mu\otimes\nu)(E) = 0$. Since

$$\{h_x \ne 0\} = \{y \in Y : h(x,y) \ne 0\} \subset {}_xE \text{ and}$$
$$\{h_y \ne 0\} = \{x \in X : h(x,y) \ne 0\} \subset E_y$$

we learn that for μ a.e. x and ν a.e. y that $\{h_x \ne 0\} \in \mathcal{M}$, $\{h_y \ne 0\} \in \mathcal{N}$, $\nu(\{h_x \ne 0\}) = 0$ and a.e. and $\mu(\{h_y \ne 0\}) = 0$. This implies $\int_Y h(x,y)d\nu(y)$

exists and equals 0 for μ a.e. x and similarly that $\int_X h(x,y)d\mu(x)$ exists and equals 0 for ν a.e. y. Therefore

$$0 = \int_{X\times Y} hd\lambda = \int_Y \left(\int_X hd\mu\right) d\nu = \int_X \left(\int_Y hd\nu\right) d\mu.$$

For general $f \in L^1(\lambda)$, we may choose $g \in L^1(\mathcal{M}\otimes\mathcal{N}, \mu\otimes\nu)$ such that $f(x,y) = g(x,y)$ for $\lambda-$ a.e. (x,y). Define $h := f - g$. Then $h = 0,\ \lambda-$ a.e. Hence by what we have just proved and Theorem 9.6 $f = g + h$ has the following properties:

1. For μ a.e. x, $y \to f(x,y) = g(x,y) + h(x,y)$ is in $L^1(\nu)$ and
$$\int_Y f(x,y)d\nu(y) = \int_Y g(x,y)d\nu(y).$$
2. For ν a.e. y, $x \to f(x,y) = g(x,y) + h(x,y)$ is in $L^1(\mu)$ and
$$\int_X f(x,y)d\mu(x) = \int_X g(x,y)d\mu(x).$$

From these assertions and Theorem 9.6, it follows that

$$\begin{aligned}\int_X d\mu(x)\int_Y d\nu(y)f(x,y) &= \int_X d\mu(x)\int_Y d\nu(y)g(x,y)\\ &= \int_Y d\nu(y)\int_Y d\nu(x)g(x,y)\\ &= \int_{X\times Y} g(x,y)d(\mu\otimes\nu)(x,y)\\ &= \int_{X\times Y} f(x,y)d\lambda(x,y).\end{aligned}$$

Similarly it is shown that

$$\int_Y d\nu(y)\int_X d\mu(x)f(x,y) = \int_{X\times Y} f(x,y)d\lambda(x,y).$$

■

9.5 Lebesgue Measure on $\mathbb{R}^d$ and the Change of Variables Theorem

Notation 9.16 *Let*

$$m^d := \overbrace{m\otimes\cdots\otimes m}^{d \text{ times}} \text{ on } \mathcal{B}_{\mathbb{R}^d} = \overbrace{\mathcal{B}_{\mathbb{R}}\otimes\cdots\otimes\mathcal{B}_{\mathbb{R}}}^{d \text{ times}}$$

be the d – fold product of Lebesgue measure m on $\mathcal{B}_{\mathbb{R}}$. We will also use m^d to denote its completion and let $\mathcal{L}_d$ be the completion of $\mathcal{B}_{\mathbb{R}^d}$ relative to m^d. A subset $A \in \mathcal{L}_d$ is called a Lebesgue measurable set and m^d is called d – dimensional Lebesgue measure, or just Lebesgue measure for short.

Definition 9.17. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is* ***Lebesgue measurable*** *if $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{L}_d$.*

Notation 9.18 *I will often be sloppy in the sequel and write m for m^d and dx for $dm(x) = dm^d(x)$, i.e.*

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} fdm = \int_{\mathbb{R}^d} fdm^d.$$

Hopefully the reader will understand the meaning from the context.

Theorem 9.19. *Lebesgue measure m^d is translation invariant. Moreover m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$.*

Proof. Let $A = J_1 \times \cdots \times J_d$ with $J_i \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}^d$. Then

$$x + A = (x_1 + J_1) \times (x_2 + J_2) \times \cdots \times (x_d + J_d)$$

and therefore by translation invariance of m on $\mathcal{B}_{\mathbb{R}}$ we find that

$$m^d(x + A) = m(x_1 + J_1)\ldots m(x_d + J_d) = m(J_1)\ldots m(J_d) = m^d(A)$$

and hence $m^d(x + A) = m^d(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ since it holds for A in a multiplicative system which generates $\mathcal{B}_{\mathbb{R}^d}$. From this fact we see that the measure $m^d(x+\cdot)$ and $m^d(\cdot)$ have the same null sets. Using this it is easily seen that $m(x + A) = m(A)$ for all $A \in \mathcal{L}_d$. The proof of the second assertion is Exercise 9.6. ■

Exercise 9.1. In this problem you are asked to show there is no reasonable notion of Lebesgue measure on an infinite dimensional Hilbert space. To be more precise, suppose H is an infinite dimensional Hilbert space and m is a **countably additive** measure on $\mathcal{B}_H$ which is invariant under translations and satisfies, $m(B_0(\varepsilon)) > 0$ for all $\varepsilon > 0$. Show $m(V) = \infty$ for all non-empty open subsets $V \subset H$.

Theorem 9.20 (Change of Variables Theorem). *Let $\Omega \subset_o \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ be a C^1 – diffeomorphism,*[2] *see Figure 9.1. Then for any Borel measurable function, $f : T(\Omega) \to [0,\infty]$,*

[2] That is $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ is a continuously differentiable bijection and the inverse map $T^{-1} : T(\Omega) \to \Omega$ is also continuously differentiable.

$$\int_{\Omega} f(T(x)) \,|\det T'(x)|\, dx = \int_{T(\Omega)} f(y)\, dy, \tag{9.24}$$

where $T'(x)$ is the linear transformation on $\mathbb{R}^d$ defined by $T'(x)v := \frac{d}{dt}|_0 T(x+tv)$. More explicitly, viewing vectors in $\mathbb{R}^d$ as columns, $T'(x)$ may be represented by the matrix

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \dots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \dots & \partial_d T_d(x) \end{bmatrix}, \tag{9.25}$$

i.e. the i - j – matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \dots, T_d(x))^{\mathrm{tr}}$ and $\partial_i = \partial/\partial x_i$.

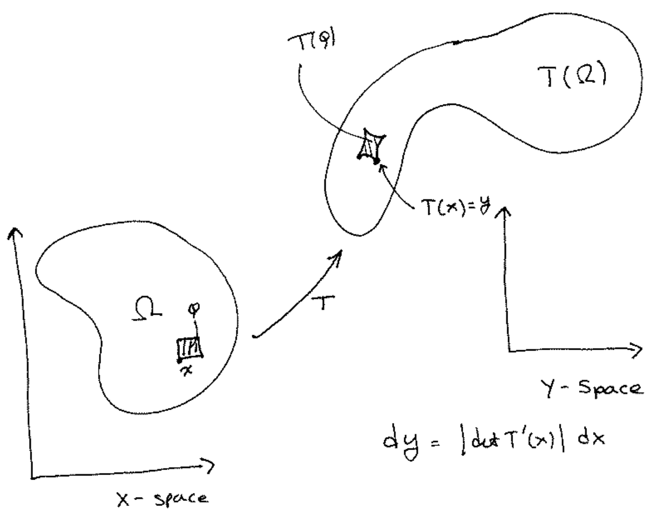


Fig. 9.1. The geometric setup of Theorem 9.20.

Remark 9.21. Theorem 9.20 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = |\det T'(x)|dx$. As usual, you must also change the limits of integration appropriately, i.e. if x ranges through Ω then y must range through $T(\Omega)$.

Note: you may skip the rest of this chapter!

Proof. The proof will be by induction on d. The case $d = 1$ was essentially done in Exercise 7.10. Nevertheless, for the sake of completeness let us give

a proof here. Suppose $d = 1$, $a < \alpha < \beta < b$ such that $[a, b]$ is a compact subinterval of Ω. Then $|\det T'| = |T'|$ and

$$\int_{[a,b]} 1_{T((\alpha,\beta])}(T(x))\, |T'(x)|\, dx = \int_{[a,b]} 1_{(\alpha,\beta]}(x)\, |T'(x)|\, dx = \int_\alpha^\beta |T'(x)|\, dx.$$

If $T'(x) > 0$ on $[a, b]$, then

$$\begin{aligned} \int_\alpha^\beta |T'(x)|\, dx &= \int_\alpha^\beta T'(x)\, dx = T(\beta) - T(\alpha) \\ &= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y)\, dy \end{aligned}$$

while if $T'(x) < 0$ on $[a, b]$, then

$$\begin{aligned} \int_\alpha^\beta |T'(x)|\, dx &= -\int_\alpha^\beta T'(x)\, dx = T(\alpha) - T(\beta) \\ &= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y)\, dy. \end{aligned}$$

Combining the previous three equations shows

$$\int_{[a,b]} f(T(x))\, |T'(x)|\, dx = \int_{T([a,b])} f(y)\, dy \tag{9.26}$$

whenever f is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 8.6 then implies that Eq. (9.26) holds for every bounded measurable function $f : T([a, b]) \to \mathbb{R}$. (Observe that $|T'(x)|$ is continuous and hence bounded for x in the compact interval, $[a, b]$.) Recall that $\Omega = \sum_{n=1}^N (a_n, b_n)$ where $a_n, b_n \in \mathbb{R} \cup \{\pm\infty\}$ for $n = 1, 2, \cdots < N$ with $N = \infty$ possible. Hence if $f : T(\Omega) \to \mathbb{R}_+$ is a Borel measurable function and $a_n < \alpha_k < \beta_k < b_n$ with $\alpha_k \downarrow a_n$ and $\beta_k \uparrow b_n$, then by what we have already proved and the monotone convergence theorem

$$\begin{aligned} \int_\Omega 1_{(a_n,b_n)} \cdot (f \circ T) \cdot |T'| dm &= \int_\Omega \left(1_{T((a_n,b_n))} \cdot f\right) \circ T \cdot |T'| dm \\ &= \lim_{k\to\infty} \int_\Omega \left(1_{T([\alpha_k,\beta_k])} \cdot f\right) \circ T \cdot |T'|\, dm \\ &= \lim_{k\to\infty} \int_{T(\Omega)} 1_{T([\alpha_k,\beta_k])} \cdot f\, dm \\ &= \int_{T(\Omega)} 1_{T((a_n,b_n))} \cdot f\, dm. \end{aligned}$$

Summing this equality on n, then shows Eq. (9.24) holds.

To carry out the induction step, we now suppose $d > 1$ and suppose the theorem is valid with d being replaced by $d-1$. For notational compactness, let us write vectors in $\mathbb{R}^d$ as row vectors rather than column vectors. Nevertheless, the matrix associated to the differential, $T'(x)$, will always be taken to be given as in Eq. (9.25).

Case 1. Suppose $T(x)$ has the form

$$T(x) = (x_i, T_2(x), \ldots, T_d(x)) \tag{9.27}$$

or

$$T(x) = (T_1(x), \ldots, T_{d-1}(x), x_i) \tag{9.28}$$

for some $i \in \{1, \ldots, d\}$. For definiteness we will assume T is as in Eq. (9.27), the case of T in Eq. (9.28) may be handled similarly. For $t \in \mathbb{R}$, let $i_t : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the inclusion map defined by

$$i_t(w) := w_t := (w_1, \ldots, w_{i-1}, t, w_{i+1}, \ldots, w_{d-1}),$$

Ω_t be the (possibly empty) open subset of $\mathbb{R}^{d-1}$ defined by

$$\Omega_t := \left\{ w \in \mathbb{R}^{d-1} : (w_1, \ldots, w_{i-1}, t, w_{i+1}, \ldots, w_{d-1}) \in \Omega \right\}$$

and $T_t : \Omega_t \to \mathbb{R}^{d-1}$ be defined by

$$T_t(w) = (T_2(w_t), \ldots, T_d(w_t)),$$

see Figure 9.2. Expanding $\det T'(w_t)$ along the first row of the matrix $T'(w_t)$ shows

$$|\det T'(w_t)| = |\det T_t'(w)|.$$

Now by the Fubini-Tonelli Theorem and the induction hypothesis,

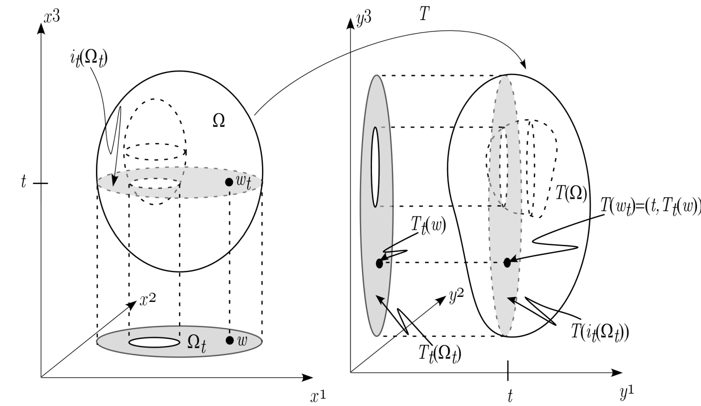


Fig. 9.2. In this picture $d = i = 3$ and Ω is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map T and slicing the set Ω along planes where $x_3 = t$.

$$\begin{aligned}
\int_\Omega f \circ T |\det T'| dm &= \int_{\mathbb{R}^d} 1_\Omega \cdot f \circ T |\det T'| dm \\
&= \int_{\mathbb{R}^d} 1_\Omega(w_t) (f \circ T)(w_t) |\det T'(w_t)| dw dt \\
&= \int_{\mathbb{R}} \left[\int_{\Omega_t} (f \circ T)(w_t) |\det T'(w_t)| dw \right] dt \\
&= \int_{\mathbb{R}} \left[\int_{\Omega_t} f(t, T_t(w)) |\det T_t'(w)| dw \right] dt \\
&= \int_{\mathbb{R}} \left[\int_{T_t(\Omega_t)} f(t, z) dz \right] dt = \int_{\mathbb{R}} \left[\int_{\mathbb{R}^{d-1}} 1_{T(\Omega)}(t, z) f(t, z) dz \right] dt \\
&= \int_{T(\Omega)} f(y) dy
\end{aligned}$$

wherein the last two equalities we have used Fubini-Tonelli along with the identity;

$$T(\Omega) = \sum_{t \in \mathbb{R}} T(i_t(\Omega)) = \sum_{t \in \mathbb{R}} \{(t, z) : z \in T_t(\Omega_t)\}.$$

Case 2. (Eq. (9.24) is true locally.) Suppose that $T : \Omega \to \mathbb{R}^d$ is a general map as in the statement of the theorem and $x_0 \in \Omega$ is an arbitrary point. We will now show there exists an open neighborhood $W \subset \Omega$ of x_0 such that

$$\int_W f \circ T |\det T'| dm = \int_{T(W)} f dm$$

holds for all Borel measurable function, $f : T(W) \to [0, \infty]$. Let M_i be the 1-i minor of $T'(x_0)$, i.e. the determinant of $T'(x_0)$ with the first row and i^{th} – column removed. Since

$$0 \neq \det T'(x_0) = \sum_{i=1}^{d} (-1)^{i+1} \partial_i T_j(x_0) \cdot M_i,$$

there must be some i such that $M_i \neq 0$. Fix an i such that $M_i \neq 0$ and let,

$$S(x) := (x_i, T_2(x), \dots, T_d(x)). \tag{9.29}$$

Observe that $|\det S'(x_0)| = |M_i| \neq 0$. Hence by the inverse function Theorem, there exist an open neighborhood W of x_0 such that $W \subset_o \Omega$ and $S(W) \subset_o \mathbb{R}^d$ and $S : W \to S(W)$ is a C^1 – diffeomorphism. Let $R : S(W) \to T(W) \subset_o \mathbb{R}^d$ to be the C^1 – diffeomorphism defined by

$$R(z) := T \circ S^{-1}(z) \text{ for all } z \in S(W).$$

Because

$$(T_1(x), \dots, T_d(x)) = T(x) = R(S(x)) = R((x_i, T_2(x), \dots, T_d(x)))$$

for all $x \in W$, if

$$(z_1, z_2, \dots, z_d) = S(x) = (x_i, T_2(x), \dots, T_d(x))$$

then

$$R(z) = \left(T_1\left(S^{-1}(z)\right), z_2, \dots, z_d\right). \tag{9.30}$$

Observe that S is a map of the form in Eq. (9.27), R is a map of the form in Eq. (9.28), $T'(x) = R'(S(x)) S'(x)$ (by the chain rule) and (by the multiplicative property of the determinant)

$$|\det T'(x)| = |\det R'(S(x))| \, |\det S'(x)| \ \forall x \in W.$$

So if $f : T(W) \to [0, \infty]$ is a Borel measurable function, two applications of the results in Case 1. shows,

$$\begin{aligned}\int_W f \circ T \cdot |\det T'| dm &= \int_W (f \circ R \cdot |\det R'|) \circ S \cdot |\det S'| \ dm \\ &= \int_{S(W)} f \circ R \cdot |\det R'| dm = \int_{R(S(W))} f dm \\ &= \int_{T(W)} f dm\end{aligned}$$

and Case 2. is proved.

Case 3. (General Case.) Let $f : \Omega \to [0, \infty]$ be a general non-negative Borel measurable function and let

$$K_n := \{x \in \Omega : \operatorname{dist}(x, \Omega^c) \geq 1/n \text{ and } |x| \leq n\}.$$

Then each K_n is a compact subset of Ω and $K_n \uparrow \Omega$ as $n \to \infty$. Using the compactness of K_n and case 2, for each $n \in \mathbb{N}$, there is a finite open cover $\mathcal{W}_n$ of K_n such that $W \subset \Omega$ and Eq. (9.24) holds with Ω replaced by W for each $W \in \mathcal{W}_n$. Let $\{W_i\}_{i=1}^\infty$ be an enumeration of $\cup_{n=1}^\infty \mathcal{W}_n$ and set $\tilde{W}_1 = W_1$ and $\tilde{W}_i := W_i \setminus (W_1 \cup \dots \cup W_{i-1})$ for all $i \geq 2$. Then $\Omega = \sum_{i=1}^\infty \tilde{W}_i$ and by repeated use of case 2.,

$$\begin{aligned}\int_\Omega f \circ T |\det T'| dm &= \sum_{i=1}^\infty \int_\Omega 1_{\tilde{W}_i} \cdot (f \circ T) \cdot |\det T'| dm \\ &= \sum_{i=1}^\infty \int_{W_i} \left[\left(1_{T(\tilde{W}_i)} f\right) \circ T\right] \cdot |\det T'| dm \\ &= \sum_{i=1}^\infty \int_{T(W_i)} 1_{T(\tilde{W}_i)} \cdot f \ dm = \sum_{i=1}^n \int_{T(\Omega)} 1_{T(\tilde{W}_i)} \cdot f \ dm \\ &= \int_{T(\Omega)} f dm.\end{aligned}$$

■

Remark 9.22. When $d = 1$, one often learns the change of variables formula as

$$\int_a^b f(T(x)) T'(x) dx = \int_{T(a)}^{T(b)} f(y) dy \tag{9.31}$$

where $f : [a, b] \to \mathbb{R}$ is a continuous function and T is C^1 – function defined in a neighborhood of $[a, b]$. If $T' > 0$ on (a, b) then $T((a, b)) = (T(a), T(b))$ and

Eq. (9.31) is implies Eq. (9.24) with $\Omega = (a,b)$. On the other hand if $T' < 0$ on (a,b) then $T((a,b)) = (T(b), T(a))$ and Eq. (9.31) is equivalent to

$$\int_{(a,b)} f(T(x))(-|T'(x)|)\,dx = -\int_{T(b)}^{T(a)} f(y)\,dy = -\int_{T((a,b))} f(y)\,dy$$

which is again implies Eq. (9.24). On the other hand Eq. (9.31) is more general than Eq. (9.24) since it does not require T to be injective. The standard proof of Eq. (9.31) is as follows. For $z \in T([a,b])$, let

$$F(z) := \int_{T(a)}^{z} f(y)\,dy.$$

Then by the chain rule and the fundamental theorem of calculus,

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_a^b F'(T(x))\,T'(x)\,dx = \int_a^b \frac{d}{dx}\left[F(T(x))\right]dx$$
$$= F(T(x))|_a^b = \int_{T(a)}^{T(b)} f(y)\,dy.$$

An application of Dynkin's multiplicative systems theorem now shows that Eq. (9.31) holds for all bounded measurable functions f on (a,b). Then by the usual truncation argument, it also holds for all positive measurable functions on (a,b).

Example 9.23. Continuing the setup in Theorem 9.20, if $A \in \mathcal{B}_\Omega$, then

$$m(T(A)) = \int_{\mathbb{R}^d} 1_{T(A)}(y)\,dy = \int_{\mathbb{R}^d} 1_{T(A)}(Tx)\,|\det T'(x)|\,dx$$
$$= \int_{\mathbb{R}^d} 1_A(x)\,|\det T'(x)|\,dx$$

wherein the second equality we have made the change of variables, $y = T(x)$. Hence we have shown

$$d(m \circ T) = |\det T'(\cdot)|\,dm.$$

In particular if $T \in GL(d,\mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d \times d$ invertible matrices, then $m \circ T = |\det T|\,m$, i.e.

$$m(T(A)) = |\det T|\,m(A) \text{ for all} A \in \mathcal{B}_{\mathbb{R}^d}. \tag{9.32}$$

This equation also shows that $m \circ T$ and m have the same null sets and hence the equality in Eq. (9.32) is valid for any $A \in \mathcal{L}_d$.

Exercise 9.2. Show that $f \in L^1\left(T(\Omega), m^d\right)$ iff

$$\int_\Omega |f \circ T|\,|\det T'|dm < \infty$$

and if $f \in L^1\left(T(\Omega), m^d\right)$, then Eq. (9.24) holds.

Example 9.24 (Polar Coordinates). Suppose $T : (0,\infty)\times(0,2\pi) \to \mathbb{R}^2$ is defined by

$$x = T(r,\theta) = (r\cos\theta, r\sin\theta),$$

i.e. we are making the change of variable,

$$x_1 = r\cos\theta \text{ and } x_2 = r\sin\theta \text{ for } 0 < r < \infty \text{ and } 0 < \theta < 2\pi.$$

In this case

$$T'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore

$$dx = |\det T'(r,\theta)|\,drd\theta = rdrd\theta.$$

Observing that

$$\mathbb{R}^2 \setminus T((0,\infty)\times(0,2\pi)) = \ell := \{(x,0) : x \geq 0\}$$

has m^2 – measure zero, it follows from the change of variables Theorem 9.20 that

$$\int_{\mathbb{R}^2} f(x)dx = \int_0^{2\pi} d\theta \int_0^\infty dr\; r \cdot f(r(\cos\theta, \sin\theta)) \tag{9.33}$$

for any Borel measurable function $f : \mathbb{R}^2 \to [0,\infty]$.

Example 9.25 (Holomorphic Change of Variables). Suppose that $f : \Omega \subset_o \mathbb{C} \cong \mathbb{R}^2 \to \mathbb{C}$ is an injective holomorphic function such that $f'(z) \neq 0$ for all $z \in \Omega$. We may express f as

$$f(x+iy) = U(x,y) + iV(x,y)$$

for all $z = x + iy \in \Omega$. Hence if we make the change of variables,

$$w = u + iv = f(x+iy) = U(x,y) + iV(x,y)$$

then

$$dudv = \left|\det\begin{bmatrix} U_x & U_y \\ V_x & V_y \end{bmatrix}\right| dxdy = |U_xV_y - U_yV_x|\,dxdy.$$

Recalling that U and V satisfy the Cauchy Riemann equations, $U_x = V_y$ and $U_y = -V_x$ with $f' = U_x + iV_x$, we learn

$$U_x V_y - U_y V_x = U_x^2 + V_x^2 = |f'|^2 .$$

Therefore

$$dudv = |f'(x+iy)|^2 \, dxdy.$$

Example 9.26. In this example we will evaluate the integral

$$I := \iint_\Omega \left(x^4 - y^4\right) dxdy$$

where

$$\Omega = \left\{(x,y) : 1 < x^2 - y^2 < 2,\ 0 < xy < 1\right\},$$

see Figure 9.3. We are going to do this by making the change of variables,

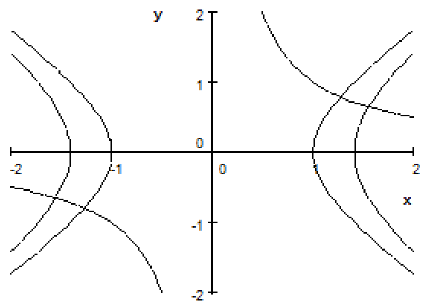


Fig. 9.3. The region Ω consists of the two curved rectangular regions shown.

$$(u,v) := T(x,y) = \left(x^2 - y^2, xy\right),$$

in which case

$$dudv = \left|\det \begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix}\right| dxdy = 2\left(x^2+y^2\right) dxdy$$

Notice that

$$\left(x^4 - y^4\right) = \left(x^2 - y^2\right)\left(x^2 + y^2\right) = u\left(x^2+y^2\right) = \frac{1}{2} ududv.$$

The function T is not injective on Ω but it is injective on each of its connected components. Let D be the connected component in the first quadrant so that

$\Omega = -D \cup D$ and $T(\pm D) = (1,2) \times (0,1)$. The change of variables theorem then implies

$$I_\pm := \iint_{\pm D} \left(x^4 - y^4\right) dxdy = \frac{1}{2} \iint_{(1,2)\times(0,1)} ududv = \frac{1}{2}\frac{u^2}{2}|_1^2 \cdot 1 = \frac{3}{4}$$

and therefore $I = I_+ + I_- = 2 \cdot (3/4) = 3/2.$

Exercise 9.3 (Spherical Coordinates). Let $T : (0,\infty)\times(0,\pi)\times(0,2\pi) \to \mathbb{R}^3$ be defined by

$$\begin{aligned} T(r,\varphi,\theta) &= (r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi) \\ &= r(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi), \end{aligned}$$

see Figure 9.4. By making the change of variables $x = T(r,\varphi,\theta)$, show

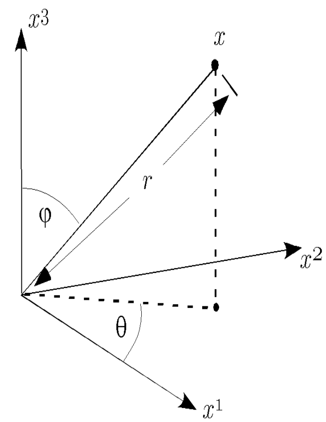


Fig. 9.4. The relation of x to (r,ϕ,θ) in spherical coordinates.

$$\int_{\mathbb{R}^3} f(x)dx = \int_0^\pi d\varphi \int_0^{2\pi} d\theta \int_0^\infty dr\ r^2 \sin\varphi \cdot f(T(r,\varphi,\theta))$$

for any Borel measurable function, $f : \mathbb{R}^3 \to [0,\infty]$.

Lemma 9.27. *Let $a > 0$ and*

$$I_d(a) := \int_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

Then $I_d(a) = (\pi/a)^{d/2}$.

Proof. By Tonelli's theorem and induction,

$$I_d(a) = \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt$$
$$= I_{d-1}(a) I_1(a) = I_1^d(a). \tag{9.34}$$

So it suffices to compute:

$$I_2(a) = \int_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)} dx_1 dx_2.$$

Using polar coordinates, see Eq. (9.33), we find,

$$I_2(a) = \int_0^\infty dr\ r \int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi \int_0^\infty re^{-ar^2} dr$$
$$= 2\pi \lim_{M\to\infty} \int_0^M re^{-ar^2} dr = 2\pi \lim_{M\to\infty} \frac{e^{-ar^2}}{-2a}\int_0^M = \frac{2\pi}{2a} = \pi/a.$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (9.34). ∎

9.6 The Polar Decomposition of Lebesgue Measure

Let

$$S^{d-1} = \{x \in \mathbb{R}^d : |x|^2 := \sum_{i=1}^d x_i^2 = 1\}$$

be the unit sphere in $\mathbb{R}^d$ equipped with its Borel σ – algebra, $\mathcal{B}_{S^{d-1}}$ and $\Phi : \mathbb{R}^d\setminus\{0\} \to (0,\infty)\times S^{d-1}$ be defined by $\Phi(x) := (|x|, |x|^{-1} x)$. The inverse map, $\Phi^{-1} : (0,\infty)\times S^{d-1} \to \mathbb{R}^d\setminus\{0\}$, is given by $\Phi^{-1}(r,\omega) = r\omega$. Since Φ and Φ^{-1} are continuous, they are both Borel measurable. For $E \in \mathcal{B}_{S^{d-1}}$ and $a > 0$, let

$$E_a := \{r\omega : r \in (0,a] \text{ and } \omega \in E\} = \Phi^{-1}((0,a]\times E) \in \mathcal{B}_{\mathbb{R}^d}.$$

Definition 9.28. *For $E \in \mathcal{B}_{S^{d-1}}$, let $\sigma(E) := d\cdot m(E_1)$. We call σ the surface measure on S^{d-1}.*

It is easy to check that σ is a measure. Indeed if $E \in \mathcal{B}_{S^{d-1}}$, then $E_1 = \Phi^{-1}((0,1]\times E) \in \mathcal{B}_{\mathbb{R}^d}$ so that $m(E_1)$ is well defined. Moreover if $E = \sum_{i=1}^\infty E_i$, then $E_1 = \sum_{i=1}^\infty (E_i)_1$ and

$$\sigma(E) = d\cdot m(E_1) = \sum_{i=1}^\infty m((E_i)_1) = \sum_{i=1}^\infty \sigma(E_i).$$

The intuition behind this definition is as follows. If $E \subset S^{d-1}$ is a set and $\varepsilon > 0$ is a small number, then the volume of

$$(1,1+\varepsilon]\cdot E = \{r\omega : r \in (1,1+\varepsilon] \text{ and } \omega \in E\}$$

should be approximately given by $m((1,1+\varepsilon]\cdot E) \cong \sigma(E)\varepsilon$, see Figure 9.5 below. On the other hand

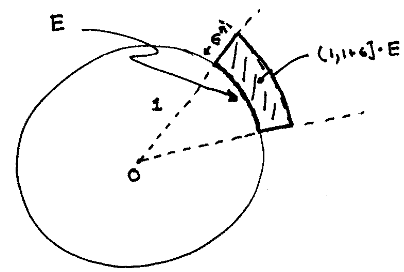


Fig. 9.5. Motivating the definition of surface measure for a sphere.

$$m((1,1+\varepsilon]E) = m(E_{1+\varepsilon}\setminus E_1) = \left\{(1+\varepsilon)^d - 1\right\} m(E_1).$$

Therefore we expect the area of E should be given by

$$\sigma(E) = \lim_{\varepsilon\downarrow 0} \frac{\left\{(1+\varepsilon)^d - 1\right\} m(E_1)}{\varepsilon} = d\cdot m(E_1).$$

The following theorem is motivated by Example 9.24 and Exercise 9.3.

Theorem 9.29 (Polar Coordinates). *If $f : \mathbb{R}^d \to [0,\infty]$ is a $(\mathcal{B}_{R^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x) dm(x) = \int_{(0,\infty)\times S^{d-1}} f(r\omega) r^{d-1}\ dr d\sigma(\omega). \tag{9.35}$$

In particular if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable then

$$\int_{\mathbb{R}^d} f(|x|) dx = \int_0^\infty f(r) dV(r) \tag{9.36}$$

where $V(r) = m(B(0,r)) = r^d m(B(0,1)) = d^{-1}\sigma\left(S^{d-1}\right) r^d$.

Proof. By Exercise 7.9,

$$\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d\setminus\{0\}} \left(f\circ\Phi^{-1}\right)\circ\Phi \; dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right) \; d\left(\Phi_* m\right) \tag{9.37}$$

and therefore to prove Eq. (9.35) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right) \;\forall\; A\in\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}. \tag{9.38}$$

If $A = (a,b]\times E$ with $0<a<b$ and $E\in\mathcal{B}_{S^{d-1}}$, then

$$\Phi^{-1}(A) = \{r\omega : r\in(a,b] \text{ and } \omega\in E\} = bE_1\setminus aE_1$$

wherein we have used $E_a = aE_1$ in the last equality. Therefore by the basic scaling properties of m and the fundamental theorem of calculus,

$$\begin{aligned}(\Phi_* m)\left((a,b]\times E\right) &= m\left(bE_1\setminus aE_1\right) = m(bE_1) - m(aE_1)\\ &= b^d m(E_1) - a^d m(E_1) = d\cdot m(E_1)\int_a^b r^{d-1}dr.\end{aligned} \tag{9.39}$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$\rho(J) = \int_J r^{d-1}dr \;\forall\; J\in\mathcal{B}_{(0,\infty)}, \tag{9.40}$$

Eq. (9.39) may be written as

$$(\Phi_* m)\left((a,b]\times E\right) = \rho((a,b])\cdot\sigma(E) = (\rho\otimes\sigma)\left((a,b]\times E\right). \tag{9.41}$$

Since

$$\mathcal{E} = \left\{(a,b]\times E : 0<a<b \text{ and } E\in\mathcal{B}_{S^{d-1}}\right\},$$

is a π class (in fact it is an elementary class) such that $\sigma(\mathcal{E}) = \mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$, it follows from the $\pi-\lambda$ Theorem and Eq. (9.41) that $\Phi_* m = \rho\otimes\sigma$. Using this result in Eq. (9.37) gives

$$\int_{\mathbb{R}^d} f dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right) \; d\left(\rho\otimes\sigma\right)$$

which combined with Tonelli's Theorem 9.6 proves Eq. (9.37). ■

Corollary 9.30. *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1}\subset\mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{9.42}$$

where Γ is the gamma function is as in Example 7.45 and 7.48.

Proof. Using Theorem 9.29 we find

$$I_d(1) = \int_0^\infty dr\; r^{d-1}e^{-r^2}\int_{S^{d-1}} d\sigma = \sigma(S^{d-1})\int_0^\infty r^{d-1}e^{-r^2}dr.$$

We simplify this last integral by making the change of variables $u = r^2$ so that $r = u^{1/2}$ and $dr = \frac{1}{2}u^{-1/2}du$. The result is

$$\begin{aligned}\int_0^\infty r^{d-1}e^{-r^2}dr &= \int_0^\infty u^{\frac{d-1}{2}}e^{-u}\frac{1}{2}u^{-1/2}du\\ &= \frac{1}{2}\int_0^\infty u^{\frac{d}{2}-1}e^{-u}du = \frac{1}{2}\Gamma(d/2).\end{aligned} \tag{9.43}$$

Combing the the last two equations with Lemma 9.27 which states that $I_d(1) = \pi^{d/2}$, we conclude that

$$\pi^{d/2} = I_d(1) = \frac{1}{2}\sigma(S^{d-1})\Gamma(d/2)$$

which proves Eq. (9.42). ■

9.7 More Spherical Coordinates

In this section we will define spherical coordinates in all dimensions. Along the way we will develop an explicit method for computing surface integrals on spheres. As usual when $n=2$ define spherical coordinates $(r,\theta)\in(0,\infty)\times[0,2\pi)$ so that

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}r\cos\theta\\ r\sin\theta\end{pmatrix} = T_2(\theta,r).$$

For $n=3$ we let $x_3 = r\cos\varphi_1$ and then

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} = T_2(\theta, r\sin\varphi_1),$$

as can be seen from Figure 9.6, so that

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}T_2(\theta,r\sin\varphi_1)\\ r\cos\varphi_1\end{pmatrix} = \begin{pmatrix}r\sin\varphi_1\cos\theta\\ r\sin\varphi_1\sin\theta\\ r\cos\varphi_1\end{pmatrix} =: T_3(\theta,\varphi_1,r,).$$

We continue to work inductively this way to define

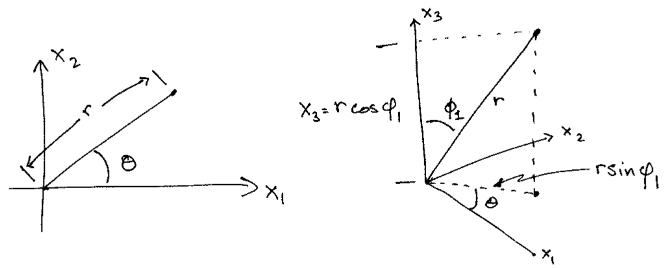


Fig. 9.6. Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r \sin \varphi_{n-1},) \\ r \cos \varphi_{n-1} \end{pmatrix} = T_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r).$$

So for example,

$$\begin{aligned} x_1 &= r \sin \varphi_2 \sin \varphi_1 \cos \theta \\ x_2 &= r \sin \varphi_2 \sin \varphi_1 \sin \theta \\ x_3 &= r \sin \varphi_2 \cos \varphi_1 \\ x_4 &= r \cos \varphi_2 \end{aligned}$$

and more generally,

$$\begin{aligned} x_1 &= r \sin \varphi_{n-2} \dots \sin \varphi_2 \sin \varphi_1 \cos \theta \\ x_2 &= r \sin \varphi_{n-2} \dots \sin \varphi_2 \sin \varphi_1 \sin \theta \\ x_3 &= r \sin \varphi_{n-2} \dots \sin \varphi_2 \cos \varphi_1 \\ &\vdots \\ x_{n-2} &= r \sin \varphi_{n-2} \sin \varphi_{n-3} \cos \varphi_{n-4} \\ x_{n-1} &= r \sin \varphi_{n-2} \cos \varphi_{n-3} \\ x_n &= r \cos \varphi_{n-2}. \end{aligned} \tag{9.44}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x) dm(x) \\ &= \int_0^\infty dr \int_{0 \le \varphi_i \le \pi, 0 \le \theta \le 2\pi} d\varphi_1 \dots d\varphi_{n-2} d\theta \begin{bmatrix} \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) \\ \times f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)) \end{bmatrix} \end{aligned} \tag{9.45}$$

where

$$\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) := \left| \det T_n'(\theta, \varphi_1, \dots, \varphi_{n-2}, r) \right|.$$

Proposition 9.31. *The Jacobian, Δ_n is given by*

$$\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) = r^{n-1} \sin^{n-2} \varphi_{n-2} \dots \sin^2 \varphi_2 \sin \varphi_1. \tag{9.46}$$

If f is a function on rS^{n-1} – the sphere of radius r centered at 0 inside of $\mathbb{R}^n$, then

$$\begin{aligned} &\int_{rS^{n-1}} f(x) d\sigma(x) = r^{n-1} \int_{S^{n-1}} f(r\omega) d\sigma(\omega) \\ &= \int_{0 \le \varphi_i \le \pi, 0 \le \theta \le 2\pi} f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)) \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) d\varphi_1 \dots d\varphi_{n-2} d\theta \end{aligned} \tag{9.47}$$

Proof. We are going to compute Δ_n inductively. Letting $\rho := r \sin \varphi_{n-1}$ and writing $\frac{\partial T_n}{\partial \xi}$ for $\frac{\partial T_n}{\partial \xi}(\theta, \varphi_1, \dots, \varphi_{n-2}, \rho)$ we have

$$\begin{aligned} &\Delta_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r) \\ &\quad = \left| \begin{bmatrix} \frac{\partial T_n}{\partial \theta} & \frac{\partial T_n}{\partial \varphi_1} & \dots & \frac{\partial T_n}{\partial \varphi_{n-2}} & \frac{\partial T_n}{\partial \rho} r \cos \varphi_{n-1} & \frac{\partial T_n}{\partial \rho} \sin \varphi_{n-1} \\ 0 & 0 & \dots & 0 & -r \sin \varphi_{n-1} & \cos \varphi_{n-1} \end{bmatrix} \right| \\ &\quad = r \left(\cos^2 \varphi_{n-1} + \sin^2 \varphi_{n-1} \right) \Delta_n(, \theta, \varphi_1, \dots, \varphi_{n-2}, \rho) \\ &\quad = r \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r \sin \varphi_{n-1}), \end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r) = r \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r \sin \varphi_{n-1}). \tag{9.48}$$

To arrive at this result we have expanded the determinant along the bottom row. Staring with $\Delta_2(\theta, r) = r$ already derived in Example 9.24, Eq. (9.48) implies,

$$\begin{aligned} \Delta_3(\theta, \varphi_1, r) &= r \Delta_2(\theta, r \sin \varphi_1) = r^2 \sin \varphi_1 \\ \Delta_4(\theta, \varphi_1, \varphi_2, r) &= r \Delta_3(\theta, \varphi_1, r \sin \varphi_2) = r^3 \sin^2 \varphi_2 \sin \varphi_1 \\ &\vdots \\ \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) &= r^{n-1} \sin^{n-2} \varphi_{n-2} \dots \sin^2 \varphi_2 \sin \varphi_1 \end{aligned}$$

which proves Eq. (9.46). Equation (9.47) now follows from Eqs. (9.35), (9.45) and (9.46). ■

As a simple application, Eq. (9.47) implies

$$\sigma(S^{n-1}) = \int_{0\le\varphi_i\le\pi,0\le\theta\le 2\pi} \sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1 d\varphi_1\dots d\varphi_{n-2}d\theta$$
$$= 2\pi\prod_{k=1}^{n-2}\gamma_k = \sigma(S^{n-2})\gamma_{n-2} \tag{9.49}$$

where $\gamma_k := \int_0^\pi \sin^k\varphi d\varphi$. If $k\ge 1$, we have by integration by parts that,

$$\gamma_k = \int_0^\pi \sin^k\varphi d\varphi = -\int_0^\pi \sin^{k-1}\varphi\ d\cos\varphi = 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \cos^2\varphi d\varphi$$
$$= 2\delta_{k,1} + (k-1)\int_0^\pi \sin^{k-2}\varphi\ \left(1-\sin^2\varphi\right)d\varphi = 2\delta_{k,1} + (k-1)\left[\gamma_{k-2}-\gamma_k\right]$$

and hence γ_k satisfies $\gamma_0 = \pi$, $\gamma_1 = 2$ and the recursion relation

$$\gamma_k = \frac{k-1}{k}\gamma_{k-2} \text{ for } k\ge 2.$$

Hence we may conclude

$$\gamma_0=\pi,\ \gamma_1=2,\ \gamma_2=\frac{1}{2}\pi,\ \gamma_3=\frac{2}{3}2,\ \gamma_4=\frac{3}{4}\frac{1}{2}\pi,\ \gamma_5=\frac{4}{5}\frac{2}{3}2,\ \gamma_6=\frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k} = \pi\frac{(2k-1)!!}{(2k)!!} \text{ and } \gamma_{2k+1} = 2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1} = \frac{2k+2}{2k+3}\gamma_{2k+1} = \frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!} = 2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)} = \frac{2k+1}{2k+1}\gamma_{2k} = \frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!} = \pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (9.49) may be written as

$$\sigma(S^n) = \sigma\left(S^{n-1}\right)\gamma_{n-1} \tag{9.50}$$

which combined with $\sigma\left(S^1\right) = 2\pi$ implies

$$\sigma\left(S^1\right) = 2\pi,$$
$$\sigma(S^2) = 2\pi\cdot\gamma_1 = 2\pi\cdot 2,$$
$$\sigma(S^3) = 2\pi\cdot 2\cdot\gamma_2 = 2\pi\cdot 2\cdot\frac{1}{2}\pi = \frac{2^2\pi^2}{2!!},$$
$$\sigma(S^4) = \frac{2^2\pi^2}{2!!}\cdot\gamma_3 = \frac{2^2\pi^2}{2!!}\cdot 2\frac{2}{3} = \frac{2^3\pi^2}{3!!}$$
$$\sigma(S^5) = 2\pi\cdot 2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi = \frac{2^3\pi^3}{4!!},$$
$$\sigma(S^6) = 2\pi\cdot 2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi\cdot\frac{4}{5}\frac{2}{3}2 = \frac{2^4\pi^3}{5!!}$$

and more generally that

$$\sigma(S^{2n}) = \frac{2\left(2\pi\right)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{\left(2\pi\right)^{n+1}}{(2n)!!} \tag{9.51}$$

which is verified inductively using Eq. (9.50). Indeed,

$$\sigma(S^{2n+1}) = \sigma(S^{2n})\gamma_{2n} = \frac{2\left(2\pi\right)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!} = \frac{\left(2\pi\right)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)}) = \sigma(S^{2n+2}) = \sigma(S^{2n+1})\gamma_{2n+1} = \frac{\left(2\pi\right)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!} = \frac{2\left(2\pi\right)^{n+1}}{(2n+1)!!}.$$

Using

$$(2n)!! = 2n\left(2(n-1)\right)\dots(2\cdot 1) = 2^n n!$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (9.35) and (9.51 are in agreement. We may also write the formula in Eq. (9.51) as

$$\sigma(S^n) = \begin{cases} \frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!} & \text{for } n \text{ odd.}\end{cases}$$

9.8 Exercises

Exercise 9.4. Prove Theorem 9.9. Suggestion, to get started define

$$\pi\left(A\right) := \int_{X_1} d\mu\left(x_1\right)\dots\int_{X_n} d\mu\left(x_n\right) 1_A\left(x_1,\dots,x_n\right)$$

and then show Eq. (9.16) holds. Use the case of two factors as the model of your proof.

Exercise 9.5. Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1, 2, 3$ be σ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1, x_2), x_3) = (x_1, x_2, x_3).$$

1. Show F is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and F^{-1} is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3,\ (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2) \times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3) \to (X_1 \times X_2 \times X_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_* [(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3] (F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then π is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0, \infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1 \times X_2 \times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1, x_2, x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

Exercise 9.6. Prove the second assertion of Theorem 9.19. That is show m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0, 1]^d) = 1$. **Hint:** Look at the proof of Theorem 5.34.

Exercise 9.7. (Part of Folland Problem 2.46 on p. 69.) Let $X = [0, 1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel σ – field on X, m be Lebesgue measure on $[0, 1]$ and ν be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x, x) \in X^2 : x \in X\}$ be the diagonal in X^2. Show

$$\int_X \left[\int_X 1_D(x, y) d\nu(y) \right] dm(x) \neq \int_X \left[\int_X 1_D(x, y) dm(x) \right] d\nu(y)$$

by explicitly computing both sides of this equation.

Exercise 9.8. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 9.9. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

Exercise 9.10. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 9.11. Folland Problem 2.56 on p. 77. Let $f \in L^1((0, a), dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0, a)$, show $g \in L^1((0, a), dm)$ and

$$\int_0^a g(x) dx = \int_0^a f(t) dt.$$

Exercise 9.12. Show $\int_0^\infty \left| \frac{\sin x}{x} \right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0, \infty), m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

Exercise 9.13. Folland Problem 2.57 on p. 77.

Exercise 9.14. Folland Problem 2.58 on p. 77.

Exercise 9.15. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 9.16. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 9.17. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 9.18. Folland Problem 2.64 on p. 80. On the integrability of $|x|^a |\log |x||^b$ for x near 0 and x near ∞ in $\mathbb{R}^n$.

Exercise 9.19. Show, using Problem 9.17 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d} \delta_{ij} \sigma\left(S^{d-1}\right).$$

Hint: show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of i and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d} \sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

References

1. Richard F. Bass, *Probabilistic techniques in analysis*, Probability and its Applications (New York), Springer-Verlag, New York, 1995. MR MR1329542 (96e:60001)
2. ______, *The Doob-Meyer decomposition revisited*, Canad. Math. Bull. **39** (1996), no. 2, 138–150. MR MR1390349 (97b:60075)
3. ______, *Diffusions and elliptic operators*, Probability and its Applications (New York), Springer-Verlag, New York, 1998. MR MR1483890 (99h:60136)
4. Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR MR1324786 (95k:60001)
5. R. M. Blumenthal and R. K. Getoor, *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York, 1968. MR MR0264757 (41 #9348)
6. K. L. Chung and R. J. Williams, *Introduction to stochastic integration*, second ed., Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1990. MR MR1102676 (92d:60057)
7. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
8. J. Dieudonné, *Foundations of modern analysis*, Academic Press, New York, 1969, Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I. MR MR0349288 (50 #1782)
9. J. L. Doob, *Stochastic processes*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR MR1038526 (91d:60002)
10. Nelson Dunford and Jacob T. Schwartz, *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. MR MR1009162 (90g:47001a)
11. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
12. ______, *Stochastic calculus*, Probability and Stochastics Series, CRC Press, Boca Raton, FL, 1996, A practical introduction. MR MR1398879 (97k:60148)
13. Evgenii B. Dynkin and Aleksandr A. Yushkevich, *Markov processes: Theorems and problems*, Translated from the Russian by James S. Wood, Plenum Press, New York, 1969. MR MR0242252 (39 #3585a)
14. Masatoshi Fukushima, Yōichi Ōshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR MR1303354 (96f:60126)
15. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
16. Paul R. Halmos, *Lectures on ergodic theory*, (1960), vii+101. MR MR0111817 (22 #2677)
17. Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam, 1989. MR MR1011252 (90m:60069)
18. Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 1987. MR MR959133 (89k:60044)

19. ______, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR MR1943877 (2003j:60001)
20. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
21. Shizuo Kakutani, *On equivalence of infinite product measures*, Ann. of Math. (2) **49** (1948), 214–224. MR MR0023331 (9,340e)
22. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
23. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR MR1121940 (92h:60127)
24. Hiroshi Kunita, *Stochastic flows and stochastic partial differential equations*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986) (Providence, RI), Amer. Math. Soc., 1987, pp. 1021–1031. MR MR934304 (89g:60194)
25. ______, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990. MR MR1070361 (91m:60107)
26. Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 463. MR 57 #1628
27. Shigeo Kusuoka and Daniel W. Stroock, *Precise asymptotics of certain Wiener functionals*, J. Funct. Anal. **99** (1991), no. 1, 1–74. MR MR1120913 (93a:60085)
28. Gregory F. Lawler, *Introduction to stochastic processes*, Chapman & Hall Probability Series, Chapman & Hall, New York, 1995. MR MR1372946 (97a:60001)
29. Terry Lyons and Zhongmin Qian, *System control and rough paths*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2002, Oxford Science Publications. MR MR2036784 (2005f:93001)
30. Zhi Ming Ma and Michael Röckner, *Introduction to the theory of (nonsymmetric) Dirichlet forms*, Universitext, Springer-Verlag, Berlin, 1992. MR MR1214375 (94d:60119)
31. Henry P. McKean, *Stochastic integrals*, AMS Chelsea Publishing, Providence, RI, 2005, Reprint of the 1969 edition, with errata. MR MR2169626 (2006d:60003)
32. Michel Métivier, *Semimartingales*, de Gruyter Studies in Mathematics, vol. 2, Walter de Gruyter & Co., Berlin, 1982, A course on stochastic processes. MR MR688144 (84i:60002)
33. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
34. James Norris, *Simplified Malliavin calculus*, Séminaire de Probabilités, XX, 1984/85, Lecture Notes in Math., vol. 1204, Springer, Berlin, 1986, pp. 101–130.
35. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
36. ______, *Brownian motion*, http://www.stat.berkeley.edu/users/peres/bmbook.pdf (2006), 1–272.
37. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
38. Philip E. Protter, *Stochastic integration and differential equations*, second ed., Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 2004, Stochastic Modelling and Applied Probability. MR MR2020294 (2005k:60008)
39. ______, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR MR2273672
40. Michael Reed and Barry Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. MR 58 #12429b
41. ______, *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Functional analysis. MR 85e:46002
42. Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999. MR MR1725357 (2000h:60050)
43. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR 2001g:60188
44. ______, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition. MR MR1780932 (2001g:60189)
45. Sheldon M. Ross, *Stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1983, Lectures in Mathematics, 14. MR MR683455 (84m:60001)
46. H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR MR1013117 (90g:00004)
47. Robert Schatten, *Norm ideals of completely continuous operators*, Second printing. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 27, Springer-Verlag, Berlin, 1970. MR 41 #2449
48. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)
49. Barry Simon, *Trace ideals and their applications*, London Mathematical Society Lecture Note Series, vol. 35, Cambridge University Press, Cambridge, 1979. MR 80k:47048
50. ______, *Functional integration and quantum physics*, second ed., AMS Chelsea Publishing, Providence, RI, 2005. MR MR2105995 (2005f:81003)
51. S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes in Mathematics, vol. 7, New York University Courant Institute of Mathematical Sciences, New York, 2001. MR MR1852999 (2003a:60001)
52. John B. Walsh, *An introduction to stochastic partial differential equations*, École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., vol. 1180, Springer, Berlin, 1986, pp. 265–439. MR MR876085 (88a:60114)