Bruce K. Driver

Math 280 (Probability Theory) Lecture Notes

January 4, 2010 *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. Due Wednesday, November 18, 2009

- Look at Lecture note Exercise 7.4, 7.9, 7.12, 7.17, 7.18, and 7.27.
- **Hand in** Lecture note Exercises: 7.5, 7.7, 7.8, 7.11, 7.13, 7.14, 7.16
- Look at from Resnik, p. 155–166: 6, 13, 26, 37
- **Hand in** from Resnick,p. 155–166: 7, 38

-3.7 Homework 7. Due Wednesday, November 25, 2009

- Look at Lecture note Exercise 9.12 – 9.14.
- Look at from Resnick§ 5.10: #18, 19, 20, 22, 31.
- **Hand in** Lecture note Exercises: 8.1, 8.2, 8.3, 8.4, 8.5, 9.4, 9.5, 9.6, 9.7, and 9.9.
- **Hand in** from Resnick § 5.10: #9, 29.

  See next page!

-3.8 Homework 8. Due Monday, December 7, 2009 by 11:00AM (Put under my office door if I am not in.)

- Look at Lecture note Exercise 10.1, 10.2, 10.3, 10.4, 10.6.
- Look at from Resnick § 4.5: 3, 5, 6, 8, 19, 28, 29.
- Look at from Resnick § 5.10: #6, 7, 8, 11, 13, 16, 22, 34

- **Hand in** Lecture note Exercises: 9.8, 10.5.
- **Hand in** from Resnick § 4.5: 1, 9*, 11, 18, 25. *Exercise 10.6 may be useful here.
- **Hand in** from Resnick § 5.10: #14, 26.

Math 280B Homework Problems Winter 2010

-2.1 Homework 1. Due Wednesday, January 13, 2010

- **Hand in** Lecture note Exercise 11.1, 11.2, 11.3, 11.4, 11.5, 11.6.
- Look at from Resnick § 5.10: #39

Math 280C Homework Problems Spring 2010

0

Math 286 Homework Problems Spring 2008

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 $\infty \cdot 0$.

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*

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

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup\{x_k : k \geq n\}. \tag{1.4}$$

[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 $n \geq N$.

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}^{-}.

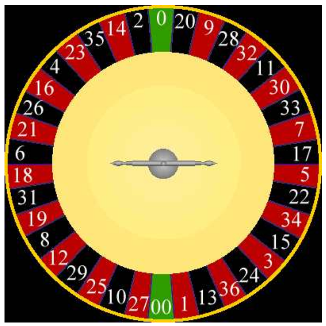
Exercise 1.2. Prove Proposition 1.10. **Hint:** apply the MCT by applying the monotone convergence theorem with $f_n(i) := \inf_{m\geq n} h_m(i)$.

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

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

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

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

we see that

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

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 $\sigma-$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

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 $X = \{1, 2, \dots, N\}$, we will write Y^N in place of $Y^{\{1,2,\dots,N\}}$ and denote $f \in Y^N$ by $f = (f_1, f_2, \dots, f_N)$ 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. $\limsup_{n\to\infty} A_n = \{x \in X : \sum_{n=1}^{\infty} 1_{A_n}(x) = \infty\}$,
4. $\liminf_{n\to\infty} A_n = \left\{x \in X : \sum_{n=1}^{\infty} 1_{A_n^c}(x) < \infty\right\}$,
5. $\sup_{k\ge n} 1_{A_k}(x) = 1_{\cup_{k\ge n} A_k} = 1_{\sup_{k\ge n} A_k}$,
6. $\inf_{k\ge n} 1_{A_k}(x) = 1_{\cap_{k\ge n} A_k} = 1_{\inf_{k\ge n} A_k}$,
7. $1_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} 1_{A_n}$, 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 $A_1, \ldots, A_n \in \mathcal{A}$ then $A_1 \cup \cdots \cup A_n \in \mathcal{A}$. 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

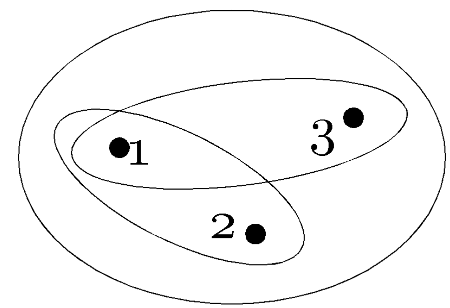


Fig. 3.1. A collection of subsets.

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

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

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

Example 3.19. Suppose that X is a set and that $\mathcal{A} \subset 2^X$ is a finite algebra, i.e. $\#(\mathcal{A}) < \infty$. 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 $\#(\mathcal{B}) = 2^n$ 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

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

wherein we have used the fact that $B_{1j_1} \cap B_{2j_2} \cap \dots \cap B_{Nj_N}$ 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

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

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

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

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 $\mathcal{S}$.

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 subadditive.*

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

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

Then

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

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,

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

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 $M:=\sup_{\Lambda\subset\subset A}\sum_{x\in\Lambda}\lambda\left(x\right)$ and for each $N\in\mathbb{N}$ let

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

Then by definition of μ,

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

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

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

from which it follows that $M\le\mu\left(A\right)$. 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 $\mu\left(\emptyset\right)=0$ 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 σ − 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 $1 \leq i \leq n$ 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 $\{B_1, \ldots, B_k\}$ 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*

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

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,

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

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. $\mathbb{E}(f) \ge 0$ for all $0 \le f \in \mathbb{S}(\mathcal{A})$. More generally, if $f, g \in \mathbb{S}(\mathcal{A})$ and $f \le g$, then $\mathbb{E}(f) \le \mathbb{E}(g)$.*
5. *For all $f \in \mathbb{S}(\mathcal{A})$,*
$$|\mathbb{E}f| \le \mathbb{E}|f|. \tag{4.19}$$

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
$$\begin{aligned}\mathbb{E}(f+g) &= \sum_{z \in \mathbb{C}} z\ P(f+g=z) \\ &= \sum_{z \in \mathbb{C}} z\ P\left(\sum_{a+b=z} \{f=a,\ g=b\}\right) \\ &= \sum_{z \in \mathbb{C}} z \sum_{a+b=z} P(\{f=a,\ g=b\}) \\ &= \sum_{z \in \mathbb{C}} \sum_{a+b=z} (a+b) P(\{f=a,\ g=b\}) \\ &= \sum_{a,b} (a+b) P(\{f=a,\ g=b\}).\end{aligned}$$
But
$$\begin{aligned}\sum_{a,b} aP(\{f=a,\ g=b\}) &= \sum_a a \sum_b P(\{f=a,\ g=b\}) \\ &= \sum_a aP(\cup_b \{f=a,\ g=b\}) \\ &= \sum_a aP(\{f=a\}) = \mathbb{E}f\end{aligned}$$
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 $f \ge 0$ then
$$\mathbb{E}(f) = \sum_{a \ge 0} aP(f=a) \ge 0$$
and if $f \le g$, then $g - f \ge 0$ so that
$$\mathbb{E}(g) - \mathbb{E}(f) = \mathbb{E}(g-f) \ge 0.$$
5. By the triangle inequality,
$$|\mathbb{E}f| = \left|\sum_{\lambda \in \mathbb{C}} \lambda P(f=\lambda)\right| \le \sum_{\lambda \in \mathbb{C}} |\lambda| P(f=\lambda) = \mathbb{E}|f|,$$
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 $\mu\left(\cup_{n=1}^M A_n\right) = \infty$,

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

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

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

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,

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

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

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

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*

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

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

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

■

4.3.2 Appendix: Riemann Stieljtes 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 $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_\mu f| \leq \mu(X)\|f\|_u. \tag{4.33}$$
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. $|\mathbb{E}_\mu f| \leq \mathbb{E}_\mu|f|$ 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 $\mathcal{S} := \{(a,b] : 0 \leq a \leq b \leq T\} \cup \{0\}$ 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 Stieljtes 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 $0 \leq a \leq b \leq T$ and $\mu_F(\{0\}) = 0$. (In fact one could allow for $\mu_F(\{0\}) = \lambda$ for any $\lambda \geq 0$, 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 $\pi := \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is a partition of $[0,T]$ and $c = (c_1,\dots,c_n) \in [0,T]^n$ with $t_{i-1} \leq c_i \leq t_i$ 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 $t_{i-1} \leq c_i \leq t_i$.

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

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

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

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

1. Explicitly compute $\int_{[0,T]} f d\mu_G$ for all $f \in C([0,1],\mathbb{C})$.
2. If $F(x) = x + \lambda \cdot 1_{x\geq a}$ 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 $i = 1, 2, \dots, n$. 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 $i = 1, 2, \dots, n$ 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, $X_1, \dots, X_n$ 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 $X_1, \dots, X_n : \Omega \to \Lambda$ and $P : 2^\Omega \to [0,1]$ be as described before Exercise 4.10. Show $X_1, \dots, X_n$ are independent iff

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

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 $i = 1, 2, \dots, n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, 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,** $\operatorname{Cov}(X,Y)$, of $X, Y \in \mathbb{S}(\mathcal{B})$ is defined by*

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

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,

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

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

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

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

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

where

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

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

$$P_x(|S_n - x| \ge \varepsilon) \le \frac{1}{4n\varepsilon^2} \quad \text{(see Figure 4.1)} \tag{4.45}$$

and hence we may conclude that

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

and therefore, that

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

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

4.4.1 Complex Weierstrass Approximation Theorem

The main goal of this subsection is to prove Theorem 4.42 which states that any continuous 2π – periodic function on $\mathbb{R}$ may be well approximated by trigonometric polynomials. The main ingredient is the following two dimensional generalization of Theorem 4.36. All of the results in this section have natural generalization to higher dimensions as well , see Theorem **??**.

Theorem 4.37 (Weierstrass Approximation Theorem). *Suppose that* $K = [0,1]^2$, $f \in C(K, \mathbb{C})$, *and*

$$p_n(x,y) := \sum_{k,l=0}^n f\left(\frac{k}{n}, \frac{l}{n}\right)\binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}. \tag{4.46}$$

Then $p_n \to f$ *uniformly on* K.

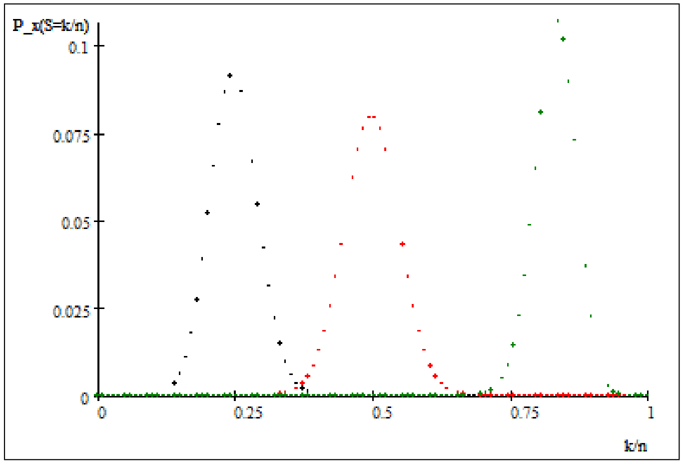


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

Proof. We are going to follow the argument given in the proof of Theorem 4.36. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([0,1]^2, \mathbb{R})$. For $(x,y) \in K$ and $n \in \mathbb{N}$ we may choose a collection of independent Bernoulli simple random variables $\{X_i, Y_i\}_{i=1}^n$ such that $P(X_i = 1) = x$ and $P(Y_i = 1) = y$ for all $1 \le i \le n$. Then letting $S_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n := \frac{1}{n}\sum_{i=1}^n Y_i$, we have

$$\mathbb{E}[f(S_n, T_n)] = \sum_{k,l=0}^{n} f\left(\frac{k}{n}, \frac{l}{n}\right) P(n \cdot S_n = k, n \cdot T_n = l) = p_n(x,y)$$

where $p_n(x,y)$ is the polynomial given in Eq. (4.46) wherein the assumed independence is needed to show,

$$P(n \cdot S_n = k, n \cdot T_n = l) = \binom{n}{k}\binom{n}{l} x^k (1-x)^{n-k} y^l (1-y)^{n-l}.$$

Thus if $M = \sup\{|f(x,y)| : (x,y) \in K\}$, $\varepsilon > 0$,

$$\delta_\varepsilon = \sup\{|f(x',y') - f(x,y)| : (x,y), (x',y') \in K \text{ and } \|x',y' - (x,y)\| \le \varepsilon\},$$

and

$$A := \{\|(S_n, T_n) - (x,y)\| > \varepsilon\},$$

we have,

$$\begin{aligned}
|f(x,y) - p_n(x,y)| &= |\mathbb{E}(f(x,y) - f((S_n, T_n)))| \\
&\le \mathbb{E}|f(x,y) - f((S_n,T_n))| \\
&= \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A] \\
&\quad + \mathbb{E}[|f(x,y) - f(S_n,T_n)| : A^c] \\
&\le 2M \cdot P(A) + \delta_\varepsilon \cdot P(A^c) \\
&\le 2M \cdot P(A) + \delta_\varepsilon.
\end{aligned} \tag{4.47}$$

To estimate $P(A)$, observe that if

$$\|(S_n, T_n) - (x,y)\|^2 = (S_n - x)^2 + (T_n - y)^2 > \varepsilon^2,$$

then either,

$$(S_n - x)^2 > \varepsilon^2/2 \text{ or } (T_n - y)^2 > \varepsilon^2/2$$

and therefore by sub-additivity and Eq. (4.45) we know

$$\begin{aligned}
P(A) &\le P\left(|S_n - x| > \varepsilon/\sqrt{2}\right) + P\left(|T_n - y| > \varepsilon/\sqrt{2}\right) \\
&\le \frac{1}{2n\varepsilon^2} + \frac{1}{2n\varepsilon^2} = \frac{1}{n\varepsilon^2}.
\end{aligned} \tag{4.48}$$

Using this estimate in Eq. (4.47) gives,

$$|f(x,y) - p_n(x,y)| \le 2M \cdot \frac{1}{n\varepsilon^2} + \delta_\varepsilon$$

and as right is independent of $(x,y) \in K$ we may conclude,

$$\limsup_{n\to\infty} \sup_{(x,y)\in K} |f(x,y) - p_n(x,y)| \le \delta_\varepsilon$$

which completes the proof since $\delta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$ because f is uniformly continuous on K. ■

Remark 4.38. We can easily improve our estimate on $P(A)$ in Eq. (4.48) by a factor of two as follows. As in the proof of Theorem 4.36,

$$\begin{aligned}
\mathbb{E}\left[\|(S_n, T_n) - (x,y)\|^2\right] &= \mathbb{E}\left[(S_n - x)^2 + (T_n - y)^2\right] \\
&= \mathrm{Var}(S_n) + \mathrm{Var}(T_n) \\
&= \frac{1}{n}x(1-x) + y(1-y) \le \frac{1}{2n}.
\end{aligned}$$

Therefore by Chebyshev's inequality,

$$P(A) = P(\|(S_n,T_n) - (x,y)\| > \varepsilon) \le \frac{1}{\varepsilon^2}\mathbb{E}\|(S_n,T_n) - (x,y)\|^2 \le \frac{1}{2n\varepsilon^2}.$$

Corollary 4.39. *Suppose that $K = [a,b]\times[c,d]$ is any compact rectangle in $\mathbb{R}^2$. Then every function, $f \in C(K,\mathbb{C})$, may be uniformly approximated by polynomial functions in $(x,y) \in \mathbb{R}^2$.*

Proof. Let $F(x,y) := f(a + x(b-a), c + y(d-c))$ – a continuous function of $(x,y) \in [0,1]^2$. Given $\varepsilon > 0$, we may use Theorem Theorem 4.37 to find a polynomial, $p(x,y)$, such that $\sup_{(x,y)\in[0,1]^2} |F(x,y) - p(x,y)| \le \varepsilon$. Letting $\xi = a + x(b-a)$ and $\eta := c + y(d-c)$, it now follows that

$$\sup_{(\xi.\eta)\in K} \left| f(\xi,\eta) - p\left(\frac{\xi-a}{b-a}, \frac{\eta-c}{d-c}\right)\right| \le \varepsilon$$

which completes the proof since $p\left(\frac{\xi-a}{b-a}, \frac{\eta-c}{d-c}\right)$ is a polynomial in (ξ,η). ■

Here is a version of the complex Weierstrass approximation theorem.

Theorem 4.40 (Complex Weierstrass Approximation Theorem). *Suppose that $K \subset \mathbb{C}$ is a compact rectangle. Then there exists polynomials in $(z = x+iy, \bar{z} = x - iy)$, $p_n(z,\bar{z})$ for $z \in \mathbb{C}$, such that $\sup_{z\in K} |q_n(z,\bar{z}) - f(z)| \to 0$ as $n \to \infty$ for every $f \in C(K,\mathbb{C})$.*

Proof. The mapping $(x,y) \in \mathbb{R}\times\mathbb{R} \to z = x+iy \in \mathbb{C}$ is an isomorphism of vector spaces. Letting $\bar{z} = x - iy$ as usual, we have $x = \frac{z+\bar{z}}{2}$ and $y = \frac{z-\bar{z}}{2i}$. Therefore under this identification any polynomial $p(x,y)$ on $\mathbb{R}\times\mathbb{R}$ may be written as a polynomial q in $(z,\bar{z})$, namely

$$q(z,\bar{z}) = p\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2i}\right).$$

Conversely a polynomial q in $(z,\bar{z})$ may be thought of as a polynomial p in (x,y), namely $p(x,y) = q(x+iy, x-iy)$. Hence the result now follows from Theorem 4.37. ■

Example 4.41. Let $K = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ and $\mathcal{A}$ be the set of polynomials in $(z,\bar{z})$ restricted to S^1. Then $\mathcal{A}$ is dense in $C(S^1)$. To prove this first observe if $f \in C(S^1)$ then $F(z) = |z| f\left(\frac{z}{|z|}\right)$ for $z \ne 0$ and $F(0) = 0$ defines $F \in C(\mathbb{C})$ such that $F|_{S^1} = f$. By applying Theorem 4.40 to F restricted to a compact rectangle containing S^1 we may find $q_n(z,\bar{z})$ converging uniformly to F on K and hence on S^1. Since $\bar{z}$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$.

Theorem 4.42 (Density of Trigonometric Polynomials). *Any 2π – periodic continuous function, $f : \mathbb{R} \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form*

$$p(x) = \sum_{\lambda\in\Lambda} a_\lambda e^{i\lambda\cdot x}$$

where Λ is a finite subset of $\mathbb{Z}$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$.

Proof. For $z \in S^1$, define $F(z) := f(\theta)$ where $\theta \in \mathbb{R}$ is chosen so that $z = e^{i\theta}$. Since f is 2π – periodic, F is well defined since if θ solves $e^{i\theta} = z$ then all other solutions are of the form $\{\theta + 2\pi n : n \in \mathbb{Z}\}$. Since the map $\theta \to e^{i\theta}$ is a local homeomorphism, i.e. for any $J = (a,b)$ with $b - a < 2\pi$, the map $\theta \in J \xrightarrow{\phi} \tilde{J} := \left\{e^{i\theta} : \theta \in J\right\} \subset S^1$ is a homeomorphism, it follows that $F(z) = f \circ \phi^{-1}(z)$ for $z \in \tilde{J}$. This shows F is continuous when restricted to $\tilde{J}$. Since such sets cover S^1, it follows that F is continuous.

By Example 4.41, the polynomials in z and $\bar{z} = z^{-1}$ are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z,\bar{z}) = \sum_{0\le m,n\le N} a_{m,n} z^m \bar{z}^n$$

such that $|F(z) - p(z,\bar{z})| \le \varepsilon$ for all $z \in S^1$. Taking $z = e^{i\theta}$ then implies

$$\sup_\theta \left| f(\theta) - p\left(e^{i\theta}, e^{-i\theta}\right)\right| \le \varepsilon$$

where

$$p\left(e^{i\theta}, e^{-i\theta}\right) = \sum_{0\le m,n\le N} a_{m,n} e^{i(m-n)\theta}$$

is the desired trigonometry polynomial. ■

4.4.2 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 4.43. *Let $\mathcal{A}\odot\mathcal{B}$ be the smallest sub-algebra of $2^{X\times Y}$ containing all sets of the form $\mathcal{S} := \{A \times B : A \in \mathcal{A}$ 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;*

$$C = \sum_{i=1}^n A_i \times B_i \text{ with } A_i \times B_i \in \mathcal{S}. \tag{4.49}$$

Theorem 4.44 (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. (4.49),

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

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 4.45. 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,

$$f\otimes g(x,y) = \left(\sum_a a1_{f=a}(x)\right)\left(\sum_b b1_{g=b}(y)\right) = \sum_{a,b} ab1_{\{f=a\}\times\{g=b\}}(x,y)$$

it follows that $f \otimes g \in \mathbb{S}(\mathcal{A} \odot \mathcal{B})$. Moreover, using Fubini's Theorem 4.44 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, $\{B_1, \ldots, B_n\}$, of Ω such that $A_\omega \in \{B_1, \ldots, B_n\}$ for all $\omega \in \Omega$.

Definition 4.46 (Conditional expectation). *Let $X : \Omega \to \mathbb{R}$ be a $\mathcal{B}$ – simple random variable, i.e. $X \in \mathbb{S}(\mathcal{B})$, and*

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)}\mathbb{E}[1_{A_\omega} X] \ \ \text{for all } \omega \in \Omega, \tag{4.50}$$

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

$$\bar{X}=\sum_{i=1}^{n}\frac{\mathbb{E}[1_{B_i}X]}{P(B_i)}1_{B_i}, \tag{4.51}$$

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
$$\mathbb{E}[XZ]=\mathbb{E}[\bar{X}Z]=\mathbb{E}[\mathbb{E}_{\mathcal{A}}X\cdot Z] \tag{4.52}$$
and
$$(\mathbb{E}_{\mathcal{A}}Z)(\omega)=\begin{cases}Z(\omega)&\text{if } P(A_\omega)>0\\0&\text{if } P(A_\omega)=0\end{cases}. \tag{4.53}$$
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 $Y\in\mathbb{S}(\mathcal{A})$,
$$\mathbb{E}\left[(X-\bar{X})^2\right]\le\mathbb{E}\left[(X-Y)^2\right] \tag{4.54}$$
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 4.47 (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

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

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 the $\{y_n\}$ sequence of results are independent of the $\{x_n\}$ sequence, we should expect by the usual mantra[2] that

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

[2] That is it should not matter which sequence of independent experiments are used to compute the time averages.

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;

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \mathbb{E}[g(Y)] \text{ for all } (x,y) \in \Omega. \tag{4.55}$$

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. (4.55) 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 4.48 (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 4.47 above. Assume that $P : \mathcal{B} \to [0,1]$ is a probability measure. If X and Y are P – independent, then Eq. (4.55) 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 4.49 (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 probability measures with desirable properties. The main results of this chapter may are summarized in the following theorem.

Theorem 5.2. *A finitely additive probability measure P on an algebra, $\mathcal{A} \subset 2^{\Omega}$, 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 $A_n' := A_n \setminus (A_1 \cup \dots A_{n-1}) \in \mathcal{A}$. Then $\cup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} A_n'$ and therefore by the countable additivity and monotonicity of μ we have,

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

■

Let us now specialize to the case where $\Omega = \mathbb{R}$ and $\mathcal{A} = \mathcal{A}(\{(a,b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\})$. 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*

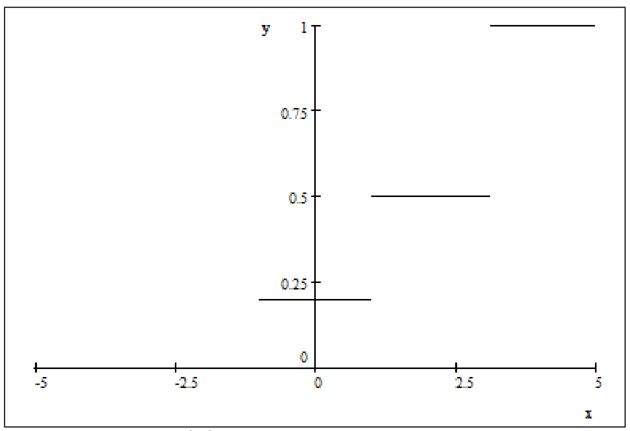
$$F(x) := P((-\infty, x]). \tag{5.1}$$

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,

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



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,

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

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

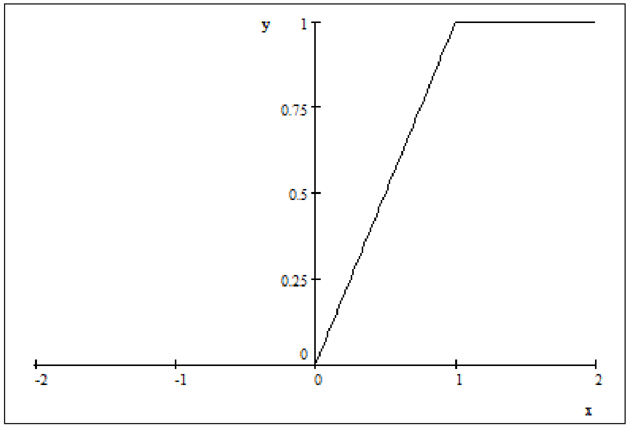


Fig. 5.1. The cumulative distribution function for the uniform distribution.

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 in Eq. (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,

$$\begin{aligned}\mu(J\cap(\alpha,b]) &= \mu(J\cap(\alpha,\bar{a}])+\mu(J\cap(\bar{a},b])\\ &\le \mu(J_l\cap(\alpha,\bar{a}])+\sum_{n=1}^{\infty}\mu(J_n\cap(\bar{a},b])\\ &= \mu(J_l\cap(\alpha,\bar{a}])+\mu\left(J_l\cap(\bar{a},b]\right)+\sum_{n\neq l}\mu(J_n\cap(\bar{a},b])\\ &= \mu(J_l\cap(\alpha,b])+\sum_{n\neq l}\mu(J_n\cap(\bar{a},b])\\ &\le \sum_{n=1}^{\infty}\mu(J_n\cap(\alpha,b]).\end{aligned}$$

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,

$$F(M)-F(a)=\mu(I_M)\le\sum_{n=1}^{\infty}\mu(J_n\cap I_M)\le\sum_{n=1}^{\infty}\mu(J_n).$$

Now let $M\to\infty$ in this last inequality to find that

$$\mu((a,\infty))=F(\infty)-F(a)\le\sum_{n=1}^{\infty}\mu(J_n).$$

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

$$p_n(s_1, \dots, s_n) := P(\{\omega \in \Omega : \omega_1 = s_1, \dots, \omega_n = s_n\}). \tag{5.21}$$

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 $\{A_m\}_{n=1}^\infty \subset \mathcal{A}$ 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*

$$E := \{x \in \Omega : P \textit{ is false for } x\}$$

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

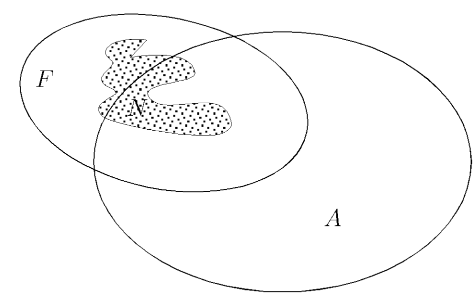


Fig. 5.2. Completing a σ – algebra.

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 $\sigma(f^{-1}(\mathcal{E})) \subset \mathcal{M}$ 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, $\pi_i(y_1,\dots,y_N) = y_i$. Then, as the reader should check,

$$\sigma(\pi_1,\dots,\pi_n) = \left\{A \times \Lambda^{N-n} : A \subset \Lambda^n\right\}.$$

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, $\mathcal{F} = \otimes_{\alpha \in I} \mathcal{F}_\alpha$.*

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

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{E}_j \text{ for all } j \in I\right\}\right)$$

and in particular,

$$\otimes_{i \in I} \mathcal{F}_i = \sigma\left(\left\{\prod_{i \in I} A_i : A_j \in \mathcal{F}_j \text{ for all } j \in I\right\}\right).$$

The last fact is easily verified directly without the aid of Exercise 6.6.

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

$$(a,b) = (a_1,b_1) \times \cdots \times (a_n,b_n). \tag{6.3}$$

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

$$A_1 \times \cdots \times A_n = \cap_{i=1}^n \pi_i^{-1}(A_i) \in \sigma(\pi_1,\dots,\pi_n) = \mathcal{B}_{\mathbb{R}^n}.$$

■

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

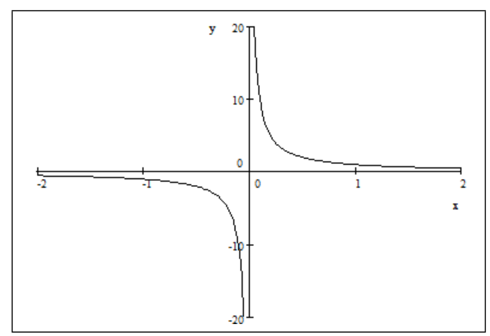
$$i(z) = \begin{cases} \frac{1}{z} & \text{if} \quad z \neq 0 \\ 0 & \text{if} \quad z = 0. \end{cases}$$

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

$$\begin{aligned} \limsup_{j\to\infty} f_j &= \inf_n \sup\{f_j : j \ge n\} \text{ and} \\ \liminf_{j\to\infty} f_j &= \sup_n \inf\{f_j : j \ge n\} \end{aligned}$$

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

$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \ne 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

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 $n\to\infty$. ■

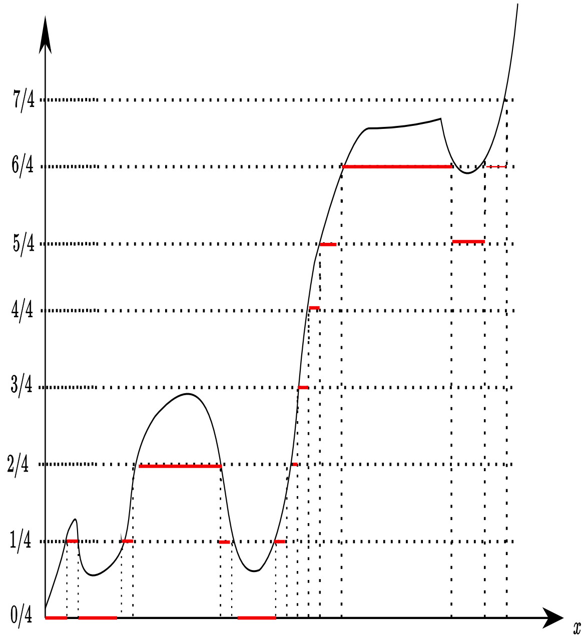
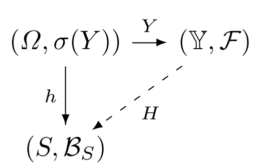


Fig. 6.1. Constructing the simple function, φ_2, approximating a function, $f : X \to [0,\infty]$. The graph of φ_2 is in red.

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 $\bar{\mathbb{R}}$.

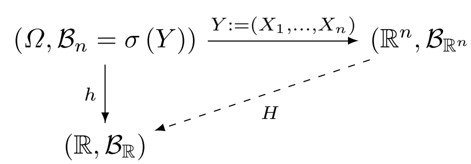
For the last assertion we may assume that $S \in \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_S = (\mathcal{B}_{\mathbb{R}})_S = \{A \cap S : A \in \mathcal{B}_{\mathbb{R}}\}$. Since $i_S : S \to \mathbb{R}$ is measurable, what we have just proved shows there exists, $H : \mathbb{Y} \to \bar{\mathbb{R}}$ 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

$$H_S = \begin{cases} H|_{H^{-1}(S)} & \text{on } \quad H^{-1}(S) \\ * & \text{on } \mathbb{Y} \setminus H^{-1}(S) \end{cases}$$

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 9.10.

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 $I = \{1, 2, \ldots, n\}$, 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 $\mathcal{B}_n := \sigma(X_1, \ldots, X_n)$, then $h : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ – measurable iff $h = H(X_1, \ldots, X_n)$ 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 $\mu \circ f^{-1}$ or $\mathrm{Law}_\mu(f)$ 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*

$$P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) \text{ for all } A_i \in \mathcal{B}_\mathbb{R}$$

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

$$\mathcal{P} := \{A_1 \times \dots \times A_n \in \mathcal{B}_{\mathbb{R}^n} : A_i \in \mathcal{B}_\mathbb{R}\}$$

for the first case and

$$\mathcal{P} := \{(-\infty, x_1] \times \dots \times (-\infty, x_n] \in \mathcal{B}_{\mathbb{R}^n} : x_i \in \mathbb{R}\}$$

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 $(X_1, \dots, X_n) \overset{d}{=} (Y_1, \dots, Y_n)$ if $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the **same distribution / law**, 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 $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ iff $(X_1, \dots, X_n) \overset{d}{=} (Y_1, \dots, Y_n)$ 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 $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$ iff $X_*P = Y_*P'$ as measures on $(\mathbb{R}^\mathbb{N}, \mathcal{F})$.

Proof. Let

$$\mathcal{P} := \cup_{n=1}^\infty \left\{A_1 \times A_2 \times \dots \times A_n \times \mathbb{R}^\mathbb{N} : A_i \in \mathcal{B}_\mathbb{R} \text{ for } 1 \le i \le n\right\}.$$

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 $A_1 \times A_2 \times \dots \times A_n \times \mathbb{R}^\mathbb{N} \in \mathcal{P}$ we have,

$$X_*P\left(A_1 \times A_2 \times \dots \times A_n \times \mathbb{R}^\mathbb{N}\right) = P((X_1, \dots, X_n) \in A_1 \times A_2 \times \dots \times A_n)$$

and hence the condition becomes,

$$P((X_1, \dots, X_n) \in A_1 \times A_2 \times \dots \times A_n) = P'((Y_1, \dots, Y_n) \in A_1 \times A_2 \times \dots \times A_n)$$

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 $(X_1, \dots, X_n) \overset{d}{=} (Y_1, \dots, Y_n)$ for all $n \in \mathbb{N}$. ■

Corollary 6.47. *Continue the notation above and assume that $\{X_i\}_{i=1}^\infty \overset{d}{=} \{Y_i\}_{i=1}^\infty$. 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 $(X_-, X_+) \overset{d}{=} (Y_-, Y_+)$ 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 $(\Omega', \mathcal{B}', P') = (\mathbb{R}^{\mathbb{N}}, \mathcal{F}, P' := X_* P)$ 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 $\{X_i\}_{i=1}^\infty \overset{\mathrm{d}}{=} \{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n := X_1 + \dots + X_n$ and $T_n := Y_1 + \dots + Y_n$. 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 $f(X_1, \dots, X_n) \overset{\mathrm{d}}{=} f(Y_1, \dots, Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty \overset{\mathrm{d}}{=} \{T_n\}_{n=1}^\infty$.
   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 $-\infty < a < b < \infty$.

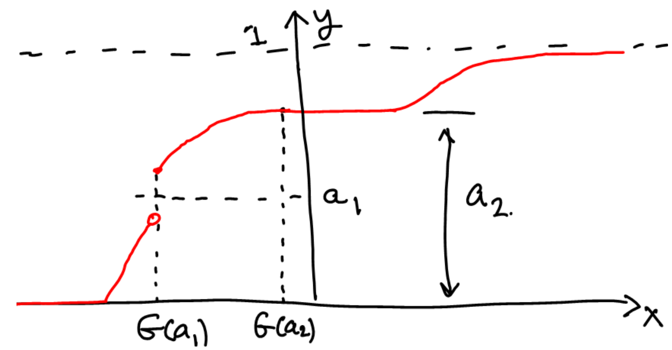


Fig. 6.2. A pictorial definition of G.

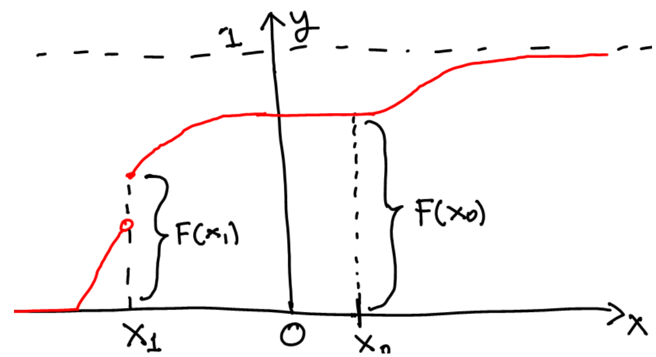


Fig. 6.3. As can be seen from this picture, $G(y) \leq x_0$ iff $y \leq F(x_0)$ and similarly, $G(y) \leq x_1$ iff $y \leq x_1$.

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* $-\infty < a < b < \infty$.

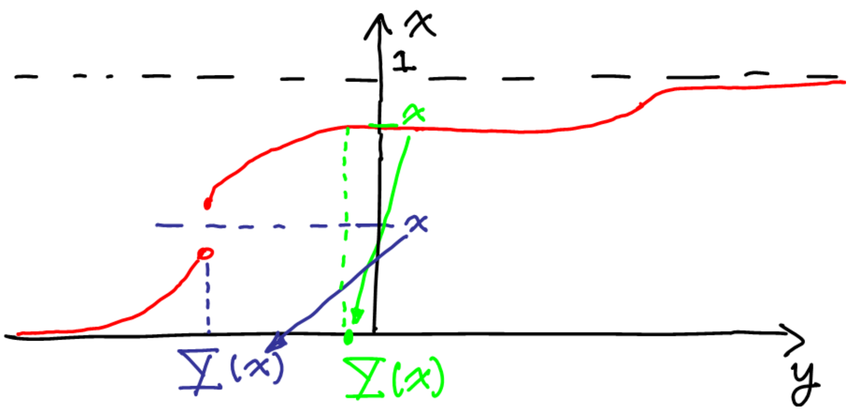


Fig. 6.4. A pictorial definition of $Y(x)$.

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 $\varphi : \Omega \to [0, \infty)$ was a measurable simple function, then we let

$$\mathbb{E}_\mu \varphi := \sum_{\lambda \in [0,\infty)} \lambda \mu \left(\varphi = \lambda\right).$$

The conventions being use here is that $0 \cdot \mu\left(\varphi = 0\right) = 0$ even when $\mu\left(\varphi = 0\right) = \infty$. This convention is necessary in order to make the integral linear – at a minimum we will want $\mathbb{E}_\mu\left[0\right] = 0$. Please be careful not blindly apply the $0 \cdot \infty = 0$ convention in other circumstances.

7.1 Integrals of positive functions

Definition 7.1. *Let $L^+ = L^+(\mathcal{B}) = \{f : \Omega \to [0, \infty] : f$ is measurable$\}$. Define*

$$\int_\Omega f(\omega)\, d\mu(\omega) = \int_\Omega f d\mu := \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\right\}.$$

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,

$$\mathbb{E} f = \int_\Omega f d\mu \text{ and } \mathbb{E}\left[f : A\right] := \int_A f d\mu.$$

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,

$$\begin{aligned}
\int_\Omega \lambda f d\mu &= \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le \lambda f\right\} \\
&= \sup\left\{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \lambda^{-1}\varphi \le f\right\} \\
&= \sup\left\{\mathbb{E}_\mu\left[\lambda \psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \sup\left\{\lambda \mathbb{E}_\mu\left[\psi\right] : \psi \text{ is simple and } \psi \le f\right\} \\
&= \lambda \int_\Omega f d\mu.
\end{aligned}$$

2. Since

$$\left\{\varphi \text{ is simple and } \varphi \le f\right\} \subset \left\{\varphi \text{ is simple and } \varphi \le g\right\},$$

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

$$1_{\{f \ge \varepsilon\}} \le 1_{\{f \ge \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \le \left(\frac{1}{\varepsilon} f\right)^p$$

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \ge \varepsilon) = \int_\Omega 1_{\{f \ge \varepsilon\}} d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega 1_{\{f \ge \varepsilon\}} f^p d\mu \le \left(\frac{1}{\varepsilon}\right)^p \int_\Omega f^p d\mu.$$

4. If $\mu(f=\infty)>0$, then $\varphi_n := n1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \le f$ for all n and hence

$$n\mu(f=\infty)=\mathbb{E}_\mu(\varphi_n)\le\int_\Omega fd\mu$$

for all n. Letting $n\to\infty$ shows $\int_\Omega fd\mu=\infty$. Thus if $\int_\Omega fd\mu<\infty$ then $\mu(f=\infty)=0$.

Moreover,

$$\{f>0\}=\cup_{n=1}^\infty\{f>1/n\}$$

with $\mu(f>1/n)\le n\int_\Omega fd\mu<\infty$ 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 $f_n\le f_m\le f$, for all $n\le m<\infty$,

$$\int f_n\le\int f_m\le\int f$$

from which if follows $\int f_n$ is increasing in n and

$$\lim_{n\to\infty}\int f_n\le\int f. \tag{7.3}$$

For the opposite inequality, let $\varphi:\Omega\to[0,\infty)$ be a simple function such that $0\le\varphi\le f$, $\alpha\in(0,1)$ and $\Omega_n:=\{f_n\ge\alpha\varphi\}$. Notice that $\Omega_n\uparrow\Omega$ and $f_n\ge\alpha 1_{\Omega_n}\varphi$ and so by definition of $\int f_n$,

$$\int f_n\ge\mathbb{E}_\mu[\alpha 1_{\Omega_n}\varphi]=\alpha\mathbb{E}_\mu[1_{\Omega_n}\varphi]. \tag{7.4}$$

Then using the identity

$$1_{\Omega_n}\varphi=1_{\Omega_n}\sum_{y>0}y1_{\{\varphi=y\}}=\sum_{y>0}y1_{\{\varphi=y\}\cap\Omega_n},$$

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 $\lim_{n\to\infty}\int f_n\ge\alpha\mathbb{E}_\mu[\varphi]$ and since $\alpha\in(0,1)$ was arbitrary we may further conclude, $\mathbb{E}_\mu[\varphi]\le\lim_{n\to\infty}\int f_n$. The latter inequality being true for all simple functions φ with $\varphi\le f$ then implies that

$$\int f=\sup_{0\le\varphi\le f}\mathbb{E}_\mu[\varphi]\le\lim_{n\to\infty}\int f_n,$$

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

$$\varphi_n:=\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}1_{\left\{\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right\}}+2^n1_{\{f>2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned}\int_\Omega fd\mu&=\lim_{n\to\infty}\int_\Omega\varphi_nd\mu\\&=\lim_{n\to\infty}\left[\sum_{k=0}^{2^{2n}-1}\frac{k}{2^n}\mu\left(\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right)+2^n\mu(f>2^n)\right].\end{aligned}$$

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 $g_N := \sum_{n=1}^{N} f_n$ and $g = \sum_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

$$\sum_{n=1}^{\infty} \int f_n := \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n$$
$$= \lim_{N\to\infty} \int g_N = \int g =: \int \sum_{n=1}^{\infty} f_n.$$

■

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

$$\sum_\Omega \varphi\rho = \sum_{\omega\in\Omega} \rho(\omega) \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}(\omega) = \sum_{z\in[0,\infty)} z \sum_{\omega\in\Omega} \rho(\omega) 1_{\{\varphi=z\}}(\omega)$$
$$= \sum_{z\in[0,\infty)} z\mu(\{\varphi=z\}) = \int_\Omega \varphi d\mu.$$

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 $f \ge 0$ is a measurable function. Then $\int_\Omega f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \ge 0$ are measurable functions such that $f \le g$ a.e. then $\int f d\mu \le \int g d\mu$. 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 Corollary and the next lemma are simple applications of Corollary 7.6.

Corollary 7.13. *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. ■

Lemma 7.14 (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 $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ■

Example 7.15. Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space (i.e. $P(\Omega) = 1$) and $X_n : \Omega \to \{0, 1\}$ are Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. If $\sum_{n=1}^{\infty} p_n < \infty$, then $P(X_n = 1 \text{ i.o.}) = 0$ and hence $P(X_n = 0 \text{ a.a.}) = 1$. In particular, $P(\lim_{n\to\infty} X_n = 0) = 1$.

7.2 Integrals of Complex Valued Functions

Definition 7.16. *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 7.17 (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 7.18. 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 7.19. *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

$$\int af = -a \int f_- + a \int f_+ = a(\int f_+ - \int f_-) = a \int f.$$

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 7.20. *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 7.21. *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 7.22. *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) *$\int_\Omega |f - g| = 0$*
   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

$$\mu(\{\mathrm{Re}(f-g)>0\})=\mu(E)=0.$$

Similar $\mu(\mathrm{Re}(f-g)<0)=0$ so that $\mathrm{Re}(f-g)=0$ a.e. Similarly, $\mathrm{Im}(f-g)=0$ a.e and hence $f-g=0$ a.e., i.e. $f=g$ a.e.

(c) $\implies$ (b) is clear and so is (b) $\implies$ (a) since

$$\left|\int_E f-\int_E g\right|\le\int|f-g|=0.$$

■

Lemma 7.23. *Suppose that $h\in\mathrm{L}^1(\mu)$ satisfies*

$$\int_A hd\mu\ge 0 \text{ for all } A\in\mathcal{B}, \tag{7.7}$$

then $h\ge 0$ a.e.

Proof. Since by assumption,

$$0=\mathrm{Im}\int_A hd\mu=\int_A \mathrm{Im}\, hd\mu \text{ for all } A\in\mathcal{B},$$

we may apply Proposition 7.22 to conclude that $\mathrm{Im}\, h=0$ a.e. Thus we may now assume that h is real valued. Taking $A=\{h<0\}$ in Eq. (7.7) implies

$$\int_\Omega 1_A|h|\,d\mu=\int_\Omega -1_A hd\mu=-\int_A hd\mu\le 0.$$

However $1_A|h|\ge 0$ and therefore it follows that $\int_\Omega 1_A|h|\,d\mu=0$ and so Proposition 7.22 implies $1_A|h|=0$ a.e. which then implies $0=\mu(A)=\mu(h<0)=0$. ■

Lemma 7.24 (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 $f\ge g$ a.e. iff*

$$\int_A fd\mu\ge\int_A gd\mu \text{ for all } A\in\mathcal{B}. \tag{7.8}$$

In particular $f=g$ a.e. iff equality holds in Eq. (7.8).

Proof. It was already shown in Proposition 7.10 that $f\ge g$ a.e. implies Eq. (7.8). For the converse assertion, let $B_n:=\{f\le n1_{\Omega_n}\}$. Then from Eq. (7.8),

$$\infty>n\mu(\Omega_n)\ge\int f1_{B_n}d\mu\ge\int g1_{B_n}d\mu$$

from which it follows that both $f1_{B_n}$ and $g1_{B_n}$ are in $\mathrm{L}^1(\mu)$ and hence $h:=f1_{B_n}-g1_{B_n}\in\mathrm{L}^1(\mu)$. Using Eq. (7.8) again we know that

$$\int_A h=\int f1_{B_n\cap A}-\int g1_{B_n\cap A}\ge 0 \text{ for all } A\in\mathcal{B}.$$

An application of Lemma 7.23 implies $h\ge 0$ a.e., i.e. $f1_{B_n}\ge g1_{B_n}$ a.e. Since $B_n\uparrow\{f<\infty\}$, we may conclude that

$$f1_{\{f<\infty\}}=\lim_{n\to\infty}f1_{B_n}\ge\lim_{n\to\infty}g1_{B_n}=g1_{\{f<\infty\}} \text{ a.e.}$$

Since $f\ge g$ whenever $f=\infty$, we have shown $f\ge g$ a.e.

If equality holds in Eq. (7.8), then we know that $g\le f$ and $f\le g$ a.e., i.e. $f=g$ a.e. ■

Notice that we can not drop the σ – finiteness assumption in Lemma 7.24. For example, let μ be the measure on $\mathcal{B}$ such that $\mu(A)=\infty$ when $A\ne\emptyset$, $g=3$, 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 7.25. *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 7.26. 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,

$$\|f+g\|_p\le\|f\|_p+\|g\|_p \text{ for all } f,g\in L^p(\mu).$$

Theorem 7.27 (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 7.21 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 7.27 to handle the general case.

Corollary 7.28. *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 7.29 (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 7.28 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 7.30 (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 $\left|\frac{\partial f}{\partial t}(t,\cdot)\right| \le g$ 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,

$$|f(t,\omega) - f(t_0,\omega)| \le g(\omega)\,|t-t_0| \text{ for all } t \in J \tag{7.9}$$

and hence

$$|f(t,\omega)| \le |f(t,\omega) - f(t_0,\omega)| + |f(t_0,\omega)| \le g(\omega)\,|t-t_0| + |f(t_0,\omega)|\,.$$

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),

$$\left|\frac{f(t,\omega)-f(t_0,\omega)}{t-t_0}\right| \le g(\omega) \text{ for all } t \in J \text{ and } \omega \in \Omega.$$

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 7.31. *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

$$f'(z) = \sum_{n=0}^\infty na_n(z-z_0)^{n-1} = \sum_{n=1}^\infty na_n(z-z_0)^{n-1}. \tag{7.10}$$

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 7.30 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

$$\begin{aligned}|g'(z,n)| &= |na_n(z-z_0)^{n-1}| \le n\,|a_n|\,\rho^{n-1}\\ &\le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}|a_n|\,r^n \le \frac{1}{r}n\left(\frac{\rho}{r}\right)^{n-1}M\end{aligned}$$

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 7.30

$$f(z) = \int_X g(z,n) d\mu(n) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

is complex differentiable with the differential given as in Eq. (7.10). ■

Definition 7.32 (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 7.33. *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 7.28 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 7.31. ■

Exercise 7.3. Let $d \in \mathbb{N}$, $\Omega = \mathbb{N}_0^d$, $\mathcal{B} = 2^{\Omega}$, $\mu : \mathcal{B} \to \mathbb{N}_0 \cup \{\infty\}$ be counting measure on Ω, and for $x \in \mathbb{R}^d$ and $\omega \in \Omega$, let $x^\omega := x_1^{\omega_1} \dots x_n^{\omega_n}$. Further suppose that $f : \Omega \to \mathbb{C}$ is function and $r_i > 0$ for $1 \le i \le d$ such that

$$\sum_{\omega \in \Omega} |f(\omega)| r^\omega = \int_\Omega |f(\omega)| r^\omega d\mu(\omega) < \infty,$$

where $r := (r_1, \dots, r_d)$. Show;

1. There is a constant, $C < \infty$ such that $|f(\omega)| \le \frac{C}{r^\omega}$ for all $\omega \in \Omega$.
2. Let
$$U := \left\{x \in \mathbb{R}^d : |x_i| < r_i \ \forall\ i\right\} \text{ and } \bar{U} = \left\{x \in \mathbb{R}^d : |x_i| \le r_i \ \forall\ i\right\}$$
Show $\sum_{\omega \in \Omega} |f(\omega) x^\omega| < \infty$ for all $x \in \bar{U}$ and the function, $F : U \to \mathbb{R}$ defined by
$$F(x) = \sum_{\omega \in \Omega} f(\omega) x^\omega \text{ is continuous on } \bar{U}.$$
3. Show, for all $x \in U$ and $1 \le i \le d$, that
$$\frac{\partial}{\partial x_i} F(x) = \sum_{\omega \in \Omega} \omega_i f(\omega) x^{\omega - e_i}$$
where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ is the i^{th} – standard basis vector on $\mathbb{R}^d$.
4. For any $\alpha \in \Omega$, let $\partial^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$ and $\alpha! := \prod_{i=1}^d \alpha_i!$ Explain why we may now conclude that
$$\partial^\alpha F(x) = \sum_{\omega \in \Omega} \alpha! f(\omega) x^{\omega - \alpha} \text{ for all } x \in U. \tag{7.13}$$
5. Conclude that $f(\alpha) = \frac{(\partial \alpha F)(0)}{\alpha!}$ for all $\alpha \in \Omega$.
6. If $g : \Omega \to \mathbb{C}$ is another function such that $\sum_{\omega \in \Omega} g(\omega) x^\omega = \sum_{\omega \in \Omega} f(\omega) x^\omega$ for x in a neighborhood of $0 \in \mathbb{R}^d$, then $g(\omega) = f(\omega)$ for all $\omega \in \Omega$.

Solution to Exercise (7.3). We take each item in turn.

1. If no such C existed, then there would exist $\omega(n) \in \Omega$ such that $|f(\omega(n))| r^{\omega(n)} \ge n$ for all $n \in \mathbb{N}$ and therefore, $\sum_{\omega \in \Omega} |f(\omega)| r^\omega \ge n$ for all $n \in \mathbb{N}$ which violates the assumption that $\sum_{\omega \in \Omega} |f(\omega)| r^\omega < \infty$.
2. If $x \in \bar{U}$, then $|x^\omega| \le r^\omega$ and therefore $\sum_{\omega \in \Omega} |f(\omega) x^\omega| \le \sum_{\omega \in \Omega} |f(\omega)| r^\omega < \infty$. The continuity of F now follows by the DCT where we can take $g(\omega) := |f(\omega)| r^\omega$ as the integrable dominating function.
3. For notational simplicity assume that $i = 1$ and let $\rho_i \in (0, r_i)$ be chosen. Then for $|x_i| < \rho_i$, we have,
$$\left|\omega_1 f(\omega) x^{\omega - e_1}\right| \le \omega_1 \rho^{\omega - e_1} \frac{C}{r^\omega} =: g(\omega)$$
where $\rho = (\rho_1, \dots, \rho_d)$. Notice that $g(\omega)$ is summable since,

$$\sum_{\omega\in\Omega} g(\omega) \le \frac{C}{\rho_1}\sum_{\omega_1=0}^{\infty}\omega_1\left(\frac{\rho_1}{r_1}\right)^{\omega_1}\cdot\prod_{i=2}^{d}\sum_{\omega_i=0}^{\infty}\left(\frac{\rho_i}{r_i}\right)^{\omega_i}$$
$$\le \frac{C}{\rho_1}\prod_{i=2}^{d}\frac{1}{1-\frac{\rho_i}{r_i}}\cdot\sum_{\omega_1=0}^{\infty}\omega_1\left(\frac{\rho_1}{r_1}\right)^{\omega_1} < \infty$$

where the last sum is finite as we saw in the proof of Corollary 7.31. Thus we may apply Corollary 7.30 in order to differentiate past the integral (= sum).

4. This is a simple matter of induction. Notice that each time we differentiate, the resulting function is still defined and differentiable on all of U.
5. Setting $x = 0$ in Eq. (7.13) shows $(\partial^\alpha F)(0) = \alpha! f(\alpha)$.
6. This follows directly from the previous item since,

$$\alpha! f(\alpha) = \partial^\alpha\left(\sum_{\omega\in\Omega} f(\omega)x^\omega\right)|_{x=0} = \partial^\alpha\left(\sum_{\omega\in\Omega} g(\omega)x^\omega\right)|_{x=0} = \alpha! g(\alpha).$$

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 $|X| \le \frac{1}{2}\left(1 + |X^2|\right)$ which shows that every square integrable random variable is also integrable.

Definition 7.34. *The* ***covariance,*** $\mathrm{Cov}(X,Y)$, *of two square integrable random variables, X and Y, is defined by*

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

where $a_X := \mathbb{E}X$ and $a_Y := \mathbb{E}Y$. The ***variance*** *of X,*

$$\mathrm{Var}(X) := \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \tag{7.14}$$

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. (7.14) that

$$\mathrm{Var}(X) \le \mathbb{E}\left[X^2\right] \text{ for all } X \in L^2(P). \tag{7.15}$$

Lemma 7.35. *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 7.4 (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 7.36 (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*

$$G_N(z) := \mathbb{E}\left[z^N\right] = \sum_{n=0}^\infty P(N=n)\,z^n \text{ for } |z| \le 1. \tag{7.16}$$

By Corollary 7.31, 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 7.37 (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 $\mathbb{E}N^2 < \infty$,

$$\operatorname{Var}(N) = G''(1) + G'(1) - \left[G'(1)\right]^2. \tag{7.17}$$

Proof. By Corollary 7.31 for $|z|<1$,

$$\begin{aligned} G_N^{(k)}(z) &= \sum_{n=0}^{\infty} P(N=n)\cdot n(n-1)\dots(n-k+1)z^{n-k} \\ &= \mathbb{E}\left[N(N-1)\dots(N-k+1)z^{N-k}\right]. \end{aligned} \tag{7.18}$$

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. (7.18) to find,

$$G^{(k)}(1) = \lim_{z\uparrow 1} G^{(k)}(z) = \mathbb{E}\left[N(N-1)\dots(N-k+1)\right].$$

■

Exercise 7.5 (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 $\operatorname{Var}(N)$.

1. Bernoulli(p) : $P(N=1)=p$ and $P(N=0)=1-p$. You should find $\mathbb{E}N = p$ and $\operatorname{Var}(N) = p-p^2$.
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 $\operatorname{Var}(N) = n\left(p-p^2\right)$.
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 $\operatorname{Var}(N) = \frac{1-p}{p^2}$.
4. Poisson(λ) : $P(N=k) = \frac{\lambda^k}{k!}e^{-\lambda}$ for all $k\in\mathbb{N}_0$. You should find $\mathbb{E}N = \lambda = \operatorname{Var}(N)$.

Exercise 7.6. Let $S_{n,p} \overset{d}{=} \text{Binomial}(n,p)$, $k\in\mathbb{N}$, $p_n = \lambda_n/n$ where $\lambda_n\to\lambda>0$ as $n\to\infty$. Show that

$$\lim_{n\to\infty} P(S_{n,p_n}=k) = \frac{\lambda^k}{k!}e^{-\lambda} = P(\text{Poisson}(\lambda)=k).$$

Thus we see that for $p = O(1/n)$ and k not too large relative to n that for large n,

$$P(\text{Binomial}(n,p)=k) \cong P(\text{Poisson}(pn)=k) = \frac{(pn)^k}{k!}e^{-pn}.$$

(We will come back to the Poisson distribution and the related Poisson process later on.)

Solution to Exercise (7.6). We have,

$$\begin{aligned} P(S_{n,p_n}=k) &= \binom{n}{k}(\lambda_n/n)^k(1-\lambda_n/n)^{n-k} \\ &= \frac{\lambda_n^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}(1-\lambda_n/n)^{n-k}. \end{aligned}$$

The result now follows since,

$$\lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{n^k} = 1$$

and

$$\begin{aligned} \lim_{n\to\infty}\ln(1-\lambda_n/n)^{n-k} &= \lim_{n\to\infty}(n-k)\ln(1-\lambda_n/n) \\ &= -\lim_{n\to\infty}\left[(n-k)\lambda_n/n\right] = -\lambda. \end{aligned}$$

7.3 Integration on $\mathbb{R}$

Notation 7.38 *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 7.39. 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

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \, dm = \int_a^b f(x) \, dx \tag{7.19}$$

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. (7.19) 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 7.68 below for a more general statement along these lines.

Theorem 7.40 (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 7.27 that F is continuous on $[a,b]$. Simple manipulations show,

$$\begin{aligned} \left|\frac{F(x+h) - F(x)}{h} - f(x)\right| &= \frac{1}{|h|} \begin{cases} \left|\int_x^{x+h} [f(y) - f(x)] \, dm(y)\right| & \text{if } h > 0 \\ \left|\int_{x+h}^x [f(y) - f(x)] \, dm(y)\right| & \text{if } h < 0 \end{cases} \\ &\le \frac{1}{|h|} \begin{cases} \int_x^{x+h} |f(y) - f(x)| \, dm(y) & \text{if } h > 0 \\ \int_{x+h}^x |f(y) - f(x)| \, dm(y) & \text{if } h < 0 \end{cases} \\ &\le \sup\{|f(y) - f(x)| : y \in [x - |h|, x + |h|]\} \end{aligned}$$

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

$$\begin{aligned} \int_a^b f(x) dm(x) &= F(b) = F(b) - F(a) \\ &= (G(b) + C) - (G(a) + C) = G(b) - G(a). \end{aligned}$$

■

We can use the above results to integrate some non-Riemann integrable functions:

Example 7.41. 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 7.39 and the fundamental theorem of calculus for Riemann integrals (or Theorem 7.40 below),

$$\begin{aligned} \int_0^\infty e^{-\lambda x} dm(x) &= \lim_{N\to\infty} \int_0^N e^{-\lambda x} dm(x) = \lim_{N\to\infty} \int_0^N e^{-\lambda x} dx \\ &= -\lim_{N\to\infty} \frac{1}{\lambda} e^{-\lambda x}|_0^N = \lambda^{-1} \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{R}} \frac{1}{1+x^2} dm(x) &= \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dm(x) = \lim_{N\to\infty} \int_{-N}^N \frac{1}{1+x^2} dx \\ &= \lim_{N\to\infty} \left[\tan^{-1}(N) - \tan^{-1}(-N)\right] = \pi. \end{aligned}$$

Let us also consider the functions x^{-p}. Using the MCT and the fundamental theorem of calculus,

$$\begin{aligned}\int_{(0,1]} \frac{1}{x^p}\, dm(x) &= \lim_{n\to\infty} \int_0^1 1_{(\frac{1}{n},1]}(x)\frac{1}{x^p} dm(x) \\ &= \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x^p} dx = \lim_{n\to\infty} \frac{x^{-p+1}}{1-p}\Big|_{1/n}^1 \\ &= \begin{cases} \frac{1}{1-p} & \text{if } p<1 \\ \infty & \text{if } p>1 \end{cases}\end{aligned}$$

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 7.7. 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 7.42 (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 7.28 since

$$\begin{aligned}\sum_{n=0}^\infty \int_\alpha^\beta |a_n|\,|x|^n\, dm(x) &\le \sum_{n=0}^\infty \left(\int_0^{|\beta|} |a_n|\,|x|^n\, dm(x) + \int_0^{|\alpha|} |a_n|\,|x|^n\, dm(x)\right) \\ &\le \sum_{n=0}^\infty |a_n| \frac{|\beta|^{n+1}+|\alpha|^{n+1}}{n+1} \le 2r\sum_{n=0}^\infty |a_n|\, r^n < \infty\end{aligned}$$

where $r=\max(|\beta|,|\alpha|)$.

Example 7.43. 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 7.40,

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{|x-r_n|}}\, dx &= \int_{r_n}^1 \frac{1}{\sqrt{x-r_n}} dx + \int_0^{r_n} \frac{1}{\sqrt{r_n-x}} dx \\ &= 2\sqrt{x-r_n}|_{r_n}^1 - 2\sqrt{r_n-x}|_0^{r_n} = 2\left(\sqrt{1-r_n}-\sqrt{r_n}\right) \\ &\le 4,\end{aligned}$$

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 7.44. The following limit holds,

$$\lim_{n\to\infty} \int_0^n \left(1-\frac{x}{n}\right)^n dm(x) = 1. \tag{7.20}$$

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[1]

$$1-x\le e^{-x} \text{ for all } x\in\mathbb{R}.$$

Therefore, for $x<n$, we have

$$0\le 1-\frac{x}{n}\le e^{-x/n} \implies \left(1-\frac{x}{n}\right)^n \le \left[e^{-x/n}\right]^n = e^{-x},$$

from which it follows that

$$0\le f_n(x)\le e^{-x} \text{ for all } x\ge 0.$$

[1] Since $y=1-x$ is the tangent line to $y=e^{-x}$ at $x=0$ and e^{-x} is convex up, it follows that $1-x\le e^{-x}$ for all $x\in\mathbb{R}$.

From Example 7.41, we know

$$\int_0^\infty e^{-x} dm(x) = 1 < \infty,$$

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. (7.20) 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 7.45. Suppose that $f_n(x) := n 1_{(0, \frac{1}{n}]}(x)$ for $n \in \mathbb{N}$. Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$ while

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x) \, dx = \lim_{n\to\infty} 1 = 1 \ne 0 = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x) \, dx.$$

The problem is that the best dominating function we can take is

$$g(x) = \sup_n f_n(x) = \sum_{n=1}^\infty n \cdot 1_{(\frac{1}{n+1}, \frac{1}{n}]}(x).$$

Notice that

$$\int_{\mathbb{R}} g(x) \, dx = \sum_{n=1}^\infty n \cdot \left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^\infty \frac{1}{n+1} = \infty.$$

Example 7.46 (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 7.47. Recall from Example 7.41 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

$$0 \le \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} = x^n e^{-\lambda x} \le C_n(\varepsilon) e^{-\varepsilon x}.$$

Using this fact, Corollary 7.30 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

$$n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x} dm(x). \tag{7.21}$$

Remark 7.48. Corollary 7.30 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 7.30 to $g(t) := \int_0^\infty 1_{x \le t} dm(x)$,

$$\dot{g}(t) = \frac{d}{dt} \int_0^\infty 1_{x \le t} dm(x) \stackrel{?}{=} \int_0^\infty \frac{\partial}{\partial t} 1_{x \le t} dm(x).$$

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 7.30 has been violated in this example.)

Exercise 7.8 (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 7.9 (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;

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\, dx = -\int_{\mathbb{R}} f(x) \cdot g'(x)\, dx. \tag{7.22}$$

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

$$\int_0^\infty f'(x) \cdot g(x)\, dx = -f(0)\, g(0) - \int_0^\infty f(x) \cdot g'(x)\, dx. \tag{7.23}$$

Outline: 1. First notice that Eq. (7.22) holds if $f(x) = 0$ for $|x| \ge N$ 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 $|x| \le 1$ and $\psi(x) = 0$ if $|x| \ge 2$. For any $\varepsilon > 0$ let $\psi_\varepsilon(x) = \psi(\varepsilon x)$ Write out the identity in Eq. (7.22) 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. (7.23).

Solution to Exercise (7.9). 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,

$$\int_{\mathbb{R}} f'(x) \cdot g(x)\, \psi_\varepsilon(x)\, dx + \int_{\mathbb{R}} f(x) \cdot g(x)\, \psi'_\varepsilon(x)\, dx = -\int_{\mathbb{R}} \psi_\varepsilon(x)\, f(x) \cdot g'(x)\, dx. \tag{7.24}$$

Since $\psi_\varepsilon(x) \to 1$ boundedly and $|\psi'_\varepsilon(x)| = \varepsilon\, |\psi'(\varepsilon x)| \le C\varepsilon$, we may use the DCT to conclude,

$$\begin{aligned}&\lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} f'(x) \cdot g(x)\, \psi_\varepsilon(x)\, dx = \int_{\mathbb{R}} f'(x) \cdot g(x)\, dx, \\ &\lim_{\varepsilon\downarrow 0} \int_{\mathbb{R}} f(x) \cdot g'(x)\, \psi_\varepsilon(x)\, dx = \int_{\mathbb{R}} f(x) \cdot g'(x)\, dx, \text{ and} \\ &\left| \int_{\mathbb{R}} f(x) \cdot g(x)\, \psi'_\varepsilon(x)\, dx \right| \le C\varepsilon \cdot \int_{\mathbb{R}} |f(x) \cdot g(x)|\, dx \to 0 \text{ as } \varepsilon \downarrow 0.\end{aligned}$$

Therefore passing to the limit as $\varepsilon \downarrow 0$ in Eq. (7.24) completes the proof of Eq. (7.22). Equation (7.23) is proved in the same way.

Definition 7.49 (Gamma Function). *The* ***Gamma function***, $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ *is defined by*

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \tag{7.25}$$

(The reader should check that $\Gamma(x) < \infty$ for all $x > 0$.)

Here are some of the more basic properties of this function.

Example 7.50 (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;

$$\Gamma(x+1) = \int_0^\infty e^{-u}\ u^{x+1}\ \frac{du}{u} = \int_0^\infty u^x \left(-\frac{d}{du}\ e^{-u}\right) du$$
$$= x\int_0^\infty u^{x-1}\ e^{-u}\ du = x\ \Gamma(x).$$

In particular, it follows from items 1. and 2. and induction that

$$\Gamma(n+1) = n! \text{ for all } n \in \mathbb{N}. \tag{7.26}$$

(Equation 7.26was also proved in Eq. (7.21).)

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 9.29) that

$$\int_{-\infty}^\infty e^{-ar^2} dr = \sqrt{\frac{\pi}{a}} \text{ for all } a > 0. \tag{7.27}$$

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 7.10 (Measures and Densities). 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

$$\int_X f d\nu = \int_X f\rho d\mu. \tag{7.28}$$

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. (7.28) still holds.

Solution to Exercise (7.10). The fact that ν is a measure follows easily from Corollary 7.6. Clearly Eq. (7.28) 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 7.51 *It is customary to informally describe ν defined in Exercise 7.10 by writing $d\nu = \rho d\mu$.*

Exercise 7.11 (Abstract Change of Variables Formula). 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

$$\int_Y g d\nu = \int_X (g\circ f)\, d\mu \tag{7.29}$$

for all measurable functions $g : Y \to [0,\infty]$. **Hint:** see the hint from Exercise 7.10.

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. (7.29) holds for all $g \in \mathrm{L}^1(\nu)$.

Example 7.52. Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\{X_i\}_{i=1}^n$ are random variables on Ω with $\nu := \mathrm{Law}_P(X_1, \dots, X_n)$, then

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g \, d\nu$$

for all $g : \mathbb{R}^n \to \mathbb{R}$ which are Borel measurable and either bounded or non-negative. This follows directly from Exercise 7.11 with $f := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ and $\mu = P$.

Remark 7.53. As a special case of Example 7.52, suppose that X is a random variable on a probability space, $(\Omega, \mathcal{B}, P)$, and $F(x) := P(X \le x)$. Then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) \, dF(x) \tag{7.30}$$

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

$$d\mu_F(x) = F'(x) \, dm(x) \tag{7.31}$$

and Eq. (7.30) may be written as

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x) \, dm(x). \tag{7.32}$$

To verify Eq. (7.31) 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 $A \in \mathcal{B}_{\mathbb{R}}$. Equation 7.32 now follows from Exercise 7.10.

Exercise 7.12. 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,

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h dm \tag{7.33}$$

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. (7.33) apply Exercise 7.11 with $g = h \circ F$ and $f = F^{-1}$.

Solution to Exercise (7.12). 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 7.11 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 7.54 (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 7.33),

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}X^n.$$

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 7.55. *A random variable $T \geq 0$ is said to be* ***exponential with parameter*** *$\lambda \in [0,\infty)$ provided, $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$. We will write $T \overset{d}{=} E(\lambda)$ 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 $P(T \in (t, t+dt)) = \lambda 1_{t\geq 0} e^{-\lambda t} dt$. Applying Corollary 7.30 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

$$\mathbb{E}T^k = \int_0^\infty \tau^k e^{-\lambda\tau}\lambda d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k \int_0^\infty e^{-\lambda\tau} d\tau = \lambda\left(-\frac{d}{d\lambda}\right)^k \lambda^{-1} = k!\lambda^{-k}. \tag{7.34}$$

In particular we see that

$$\mathrm{Var}(T) = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \tag{7.35}$$

Alternatively we may compute the moment generating function for T,

$$\begin{aligned} M_T(a) := \mathbb{E}\left[e^{aT}\right] &= \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau} d\tau \\ &= \int_0^\infty e^{a\tau}\lambda e^{-\lambda\tau} d\tau = \frac{\lambda}{\lambda - a} = \frac{1}{1 - a\lambda^{-1}} \end{aligned} \tag{7.36}$$

which is valid for $a < \lambda$. On the other hand (see Proposition 7.33), we know that

$$\mathbb{E}\left[e^{aT}\right] = \sum_{n=0}^\infty \frac{a^n}{n!}\mathbb{E}[T^n] \text{ for } |a| < \lambda. \tag{7.37}$$

Comparing this with Eq. (7.36) again shows that Eq. (7.34) is valid.

Here is yet another way to understand and generalize Eq. (7.36). We simply make the change of variables, $u = \lambda\tau$ in the integral in Eq. (7.34) 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 7.56 (Memoryless property). *A random variable, $T \in (0,\infty]$ has an exponential distribution iff it satisfies the memoryless property:*

$$P(T > s+t | T > s) = P(T > t) \text{ for all } s,t \geq 0,$$

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 [8].) 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

$$g(t+s) = g(t)g(s) \text{ for all } s,t \geq 0.$$

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 $0 \leq \lambda < \infty$.

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 $r \leq t \leq s$ we have, since g is decreasing, that

$$e^{-\lambda r} = g(r) \geq g(t) \geq g(s) = e^{-\lambda s}.$$

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 7.13 (Gamma Distributions). Let X be a positive random variable. For $k, \theta > 0$, we say that $X \overset{d}{=}\text{Gamma}(k, \theta)$ 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 7.32), $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ 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 $\text{Var}(X) = k\theta^2$. (Notice that when $k = 1$ and $\theta = \lambda^{-1}$, $X \overset{d}{=} E(\lambda)$.)

7.5.1 Normal (Gaussian) Random Variables

Definition 7.57 (Normal / Gaussian Random Variables). *A random variable, Y,* ***is normal with mean*** *μ* ***standard deviation*** *σ^2 iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{7.38}$$

We will abbreviate this by writing $Y \overset{d}{=} N(\mu, \sigma^2)$. 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. (7.38) 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 $Y \overset{d}{=} N(\mu, \sigma^2)$ 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 7.50 and Lemma 9.29.

Exercise 7.14. 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

$$\begin{aligned}\mathbb{E}[Xf(X)] &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \frac{d}{dx} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(x)\, e^{-\frac{1}{2}x^2} dx = \mathbb{E}[f'(X)].\end{aligned}$$

Example 7.58. 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 7.14,

$$\alpha_{k+1} = \mathbb{E}\left[X^{2k+1} \cdot X\right] = (2k+1)\,\alpha_k \text{ with } \alpha_0 = 1.$$

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,

$$\mathbb{E}X^{2k} = \alpha_k = (2k-1)!!, \tag{7.39}$$

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,

$$\begin{aligned}\mathbb{E}|X|^\beta &= \frac{1}{\sqrt{\pi}} \int_0^\infty (2y)^{\beta/2} e^{-y} y^{-1/2} dy \\ &= \frac{1}{\sqrt{\pi}} 2^{\beta/2} \int_0^\infty y^{(\beta+1)/2} e^{-y} y^{-1} dy = \frac{1}{\sqrt{\pi}} 2^{\beta/2} \Gamma\left(\frac{\beta+1}{2}\right).\end{aligned} \tag{7.40}$$

Exercise 7.15. Suppose that $X \stackrel{d}{=} N(0,1)$ and $\lambda \in \mathbb{R}$. Show

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = \exp\left(-\lambda^2/2\right). \tag{7.41}$$

Hint: Use Corollary 7.30 to show, $f'(\lambda) = i\mathbb{E}\left[Xe^{i\lambda X}\right]$ and then use Exercise 7.14 to see that $f'(\lambda)$ satisfies a simple ordinary differential equation.

Solution to Exercise (7.15). Using Corollary 7.30 and Exercise 7.14,

$$\begin{aligned} f'(\lambda) &= i\mathbb{E}\left[Xe^{i\lambda X}\right] = i\mathbb{E}\left[\frac{d}{dX}e^{i\lambda X}\right] \\ &= i\cdot(i\lambda)\,\mathbb{E}\left[e^{i\lambda X}\right] = -\lambda f(\lambda) \text{ with } f(0) = 1. \end{aligned}$$

Solving for the unique solution of this differential equation gives Eq. (7.41).

Exercise 7.16. Suppose that $X \stackrel{d}{=} N(0,1)$ 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 7.15 or you could use a completion of the squares argument along with the translation invariance of Lebesgue measure.)

Exercise 7.17. Use Exercise 7.16 and Proposition 7.33 to give another proof that $\mathbb{E}X^{2k} = (2k-1)!!$ when $X \stackrel{d}{=} N(0,1)$.

Exercise 7.18. Let $X \stackrel{d}{=} N(0,1)$ 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 7.59 (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 $x \geq 0$,

$$P(X \geq x) \leq \min\left(\frac{1}{2} - \frac{x}{\sqrt{2\pi}}e^{-x^2/2}, \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}\right) \leq \frac{1}{2}e^{-x^2/2}. \tag{7.42}$$

Moreover (see [10, Lemma 2.5]),

$$P(X \geq x) \geq \max\left(1 - \frac{x}{\sqrt{2\pi}}, \frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right) \tag{7.43}$$

which combined with Eq. (7.42) proves Mill's ratio (see [3]);

$$\lim_{x\to\infty}\frac{P(X \geq x)}{\frac{1}{\sqrt{2\pi}x}e^{-x^2/2}} = 1. \tag{7.44}$$

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

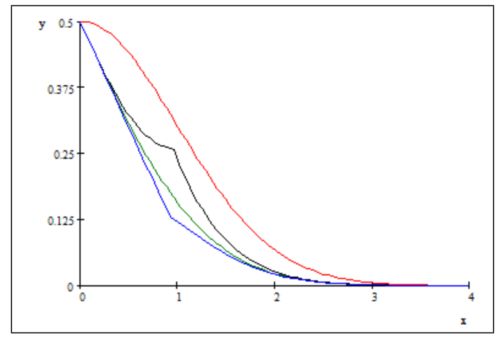


Fig. 7.1. Plots of $P(X \geq x)$ and its estimates.

Figure 7.1 is given below.

We begin by observing that

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy \leq \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{y}{x}e^{-y^2/2}dy \\ &\leq -\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-y^2/2}|_x^{-\infty} = \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}. \end{aligned} \tag{7.45}$$

If we only want to prove Mill's ratio (7.44), we could proceed as follows. Let $\alpha > 1$, then for $x > 0$,

$$\begin{aligned} P(X \geq x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy \\ &\geq \frac{1}{\sqrt{2\pi}}\int_x^{\alpha x}\frac{y}{\alpha x}e^{-y^2/2}dy = -\frac{1}{\sqrt{2\pi}}\frac{1}{\alpha x}e^{-y^2/2}|_{y=x}^{y=\alpha x} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\alpha x}e^{-x^2/2}\left[1 - e^{-\alpha^2 x^2/2}\right] \end{aligned}$$

from which it follows,

$$\liminf_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P(X\ge x)\right]\ge 1/\alpha\uparrow 1 \text{ as } \alpha\downarrow 1.$$

The estimate in Eq. (7.45) shows $\limsup_{x\to\infty}\left[\sqrt{2\pi}xe^{x^2/2}\cdot P(X\ge x)\right]\le 1.$

To get more precise estimates, we begin by observing,

$$P(X\ge x)=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy \tag{7.46}$$
$$\le\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dy\le\frac{1}{2}-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}x.$$

This equation along with Eq. (7.45) gives the first equality in Eq. (7.42). To prove the second equality observe that $\sqrt{2\pi}>2$, so

$$\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}\le\frac{1}{2}e^{-x^2/2} \text{ if } x\ge 1.$$

For $x\le 1$ we must show,

$$\frac{1}{2}-\frac{x}{\sqrt{2\pi}}e^{-x^2/2}\le\frac{1}{2}e^{-x^2/2}$$

or equivalently that $f(x):=e^{x^2/2}-\sqrt{\frac{2}{\pi}}x\le 1$ for $0\le x\le 1$. Since f is convex $\left(f''(x)=\left(x^2+1\right)e^{x^2/2}>0\right)$, $f(0)=1$ and $f(1)\cong 0.85<1$, it follows that $f\le 1$ on $[0,1]$. This proves the second inequality in Eq. (7.42).

It follows from Eq. (7.46) that

$$P(X\ge x)=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x e^{-y^2/2}dy$$
$$\ge\frac{1}{2}-\frac{1}{\sqrt{2\pi}}\int_0^x 1dy=\frac{1}{2}-\frac{1}{\sqrt{2\pi}}x \text{ for all } x\ge 0.$$

So to finish the proof of Eq. (7.43) we must show,

$$f(x):=\frac{1}{\sqrt{2\pi}}xe^{-x^2/2}-\left(1+x^2\right)P(X\ge x)$$
$$=\frac{1}{\sqrt{2\pi}}\left[xe^{-x^2/2}-\left(1+x^2\right)\int_x^\infty e^{-y^2/2}dy\right]\le 0 \text{ for all } 0\le x<\infty.$$

This follows by observing that $f(0)=-1/2<0$, $\lim_{x\uparrow\infty}f(x)=0$ and

$$f'(x)=\frac{1}{\sqrt{2\pi}}\left[e^{-x^2/2}\left(1-x^2\right)-2xP(X\ge x)+\left(1+x^2\right)e^{-x^2/2}\right]$$
$$=2\left(\frac{1}{\sqrt{2\pi}}e^{-x^2/2}-xP(X\ge y)\right)\ge 0,$$

where the last inequality is a consequence Eq. (7.42). ■

7.6 Stirling's Formula

On occasion one is faced with estimating an integral of the form, $\int_J e^{-G(t)}dt$, where $J=(a,b)\subset\mathbb{R}$ and $G(t)$ is a C^1 – function with a unique (for simplicity) global minimum at some point $t_0\in J$. The idea is that the majority contribution of the integral will often come from some neighborhood, $(t_0-\alpha,t_0+\alpha)$, of t_0. Moreover, it may happen that $G(t)$ can be well approximated on this neighborhood by its Taylor expansion to order 2;

$$G(t)\cong G(t_0)+\frac{1}{2}\ddot{G}(t_0)(t-t_0)^2.$$

Notice that the linear term is zero since t_0 is a minimum and therefore $\dot{G}(t_0)=0$. We will further assume that $\ddot{G}(t_0)\ne 0$ and hence $\ddot{G}(t_0)>0$. Under these hypothesis we will have,

$$\int_J e^{-G(t)}dt\cong e^{-G(t_0)}\int_{|t-t_0|<\alpha}\exp\left(-\frac{1}{2}\ddot{G}(t_0)(t-t_0)^2\right)dt.$$

Making the change of variables, $s=\sqrt{\ddot{G}(t_0)}(t-t_0)$, in the above integral then gives,

$$\int_J e^{-G(t)}dt\cong\frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\int_{|s|<\sqrt{\ddot{G}(t_0)}\cdot\alpha}e^{-\frac{1}{2}s^2}ds$$
$$=\frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\left[\sqrt{2\pi}-\int_{\sqrt{\ddot{G}(t_0)}\cdot\alpha}^\infty e^{-\frac{1}{2}s^2}ds\right]$$
$$=\frac{1}{\sqrt{\ddot{G}(t_0)}}e^{-G(t_0)}\left[\sqrt{2\pi}-O\left(\frac{1}{\sqrt{\ddot{G}(t_0)}\cdot\alpha}e^{-\frac{1}{2}\ddot{G}(t_0)\cdot\alpha^2}\right)\right].$$

If α is sufficiently large, for example if $\sqrt{\ddot{G}(t_0)}\cdot\alpha=3$, then the error term is about 0.0037 and we should be able to conclude that

$$\int_J e^{-G(t)}dt\cong\sqrt{\frac{2\pi}{\ddot{G}(t_0)}}e^{-G(t_0)}. \tag{7.47}$$

The proof of the next theorem (Stirling's formula for the Gamma function) will illustrate these ideas and what one has to do to carry them out rigorously.

Theorem 7.60 (Stirling's formula). *The Gamma function (see Definition 7.49), satisfies Stirling's formula,*

$$\lim_{x\to\infty}\frac{\Gamma(x+1)}{\sqrt{2\pi}e^{-x}x^{x+1/2}}=1. \tag{7.48}$$

In particular, if $n\in\mathbb{N}$, we have

$$n!=\Gamma(n+1)\sim\sqrt{2\pi}e^{-n}n^{n+1/2}$$

where we write $a_n\sim b_n$ to mean, $\lim_{n\to\infty}\frac{a_n}{b_n}=1$.

Proof. (The following proof is an elaboration of the proof found on page 236-237 in Krantz's Real Analysis and Foundations.) We begin with the formula for $\Gamma(x+1)$;

$$\Gamma(x+1)=\int_0^\infty e^{-t}t^x dt=\int_0^\infty e^{-G_x(t)}dt, \tag{7.49}$$

where

$$G_x(t):=t-x\ln t.$$

Then $\dot G_x(t)=1-x/t$, $\ddot G_x(t)=x/t^2$, G_x has a global minimum (since $\ddot G_x>0$) at $t_0=x$ where

$$G_x(x)=x-x\ln x \text{ and } \ddot G_x(x)=1/x.$$

So if Eq. (7.47) is valid in this case we should expect,

$$\Gamma(x+1)\cong\sqrt{2\pi x}e^{-(x-x\ln x)}=\sqrt{2\pi}e^{-x}x^{x+1/2}$$

which would give Stirling's formula. The rest of the proof will be spent on rigorously justifying the approximations involved.

Let us begin by making the change of variables $s=\sqrt{\ddot G(t_0)}(t-t_0)=\frac{1}{\sqrt{x}}(t-x)$ as suggested above. Then

$$\begin{aligned}G_x(t)-G_x(x)&=(t-x)-x\ln(t/x)=\sqrt{x}s-x\ln\left(\frac{x+\sqrt{x}s}{x}\right)\\&=x\left[\frac{s}{\sqrt{x}}-\ln\left(1+\frac{s}{\sqrt{x}}\right)\right]=s^2q\left(\frac{s}{\sqrt{x}}\right)\end{aligned}$$

where

$$q(u):=\frac{1}{u^2}[u-\ln(1+u)] \text{ for } u>-1 \text{ with } q(0):=\frac{1}{2}.$$

Setting $q(0)=1/2$ makes q a continuous and in fact smooth function on $(-1,\infty)$, see Figure 7.2. Using the power series expansion for $\ln(1+u)$ we find,

$$q(u)=\frac{1}{2}+\sum_{k=3}^\infty\frac{(-u)^{k-2}}{k} \text{ for } |u|<1. \tag{7.50}$$

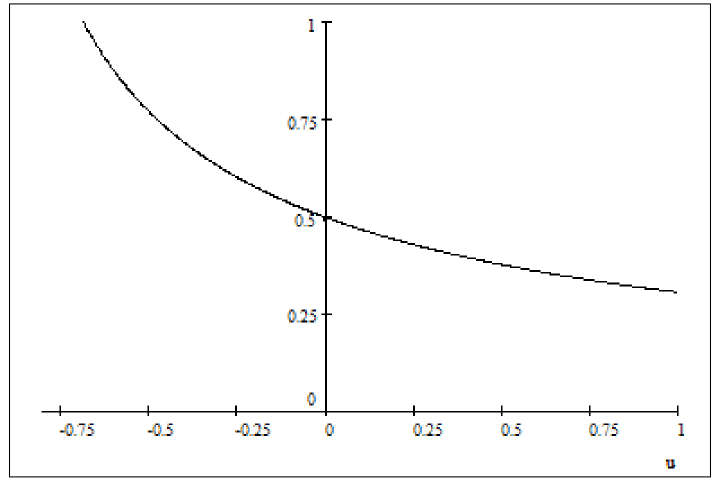


Fig. 7.2. Plot of $q(u)$.

Making the change of variables, $t=x+\sqrt{x}s$ in the second integral in Eq. (7.49) yields,

$$\Gamma(x+1)=e^{-(x-x\ln x)}\sqrt{x}\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=x^{x+1/2}e^{-x}\cdot I(x),$$

where

$$I(x)=\int_{-\sqrt{x}}^\infty e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds=\int_{-\infty}^\infty 1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}ds. \tag{7.51}$$

From Eq. (7.50) it follows that $\lim_{u\to0}q(u)=1/2$ and therefore,

$$\int_{-\infty}^\infty\lim_{x\to\infty}\left[1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\right]ds=\int_{-\infty}^\infty e^{-\frac{1}{2}s^2}ds=\sqrt{2\pi}. \tag{7.52}$$

So if there exists a dominating function, $F\in L^1(\mathbb{R},m)$, such that

$$1_{s\ge-\sqrt{x}}\cdot e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2}\le F(s) \text{ for all } s\in\mathbb{R} \text{ and } x\ge1,$$

we can apply the DCT to learn that $\lim_{x\to\infty}I(x)=\sqrt{2\pi}$ which will complete the proof of Stirling's formula.

We now construct the desired function F. From Eq. (7.50) it follows that $q(u)\ge1/2$ for $-1<u\le0$. Since $u-\ln(1+u)>0$ for $u\ne0$ ($u-\ln(1+u)$ is convex and has a minimum of 0 at $u=0$) we may conclude that $q(u)>0$ for all $u > -1$ therefore by compactness (on $[0, M]$), $\min_{-1<u\le M} q(u) = \varepsilon(M) > 0$ for all $M \in (0, \infty)$, see Remark 7.61 for more explicit estimates. Lastly, since $\frac{1}{u}\ln(1+u) \to 0$ as $u \to \infty$, there exists $M < \infty$ ($M = 3$ would due) such that $\frac{1}{u}\ln(1+u) \le \frac{1}{2}$ for $u \ge M$ and hence,

$$q(u) = \frac{1}{u}\left[1 - \frac{1}{u}\ln(1+u)\right] \ge \frac{1}{2u} \text{ for } u \ge M.$$

So there exists $\varepsilon > 0$ and $M < \infty$ such that (for all $x \ge 1$),

$$\begin{aligned} 1_{s\ge-\sqrt{x}} e^{-q\left(\frac{s}{\sqrt{x}}\right)s^2} &\le 1_{-\sqrt{x}<s\le M} e^{-\varepsilon s^2} + 1_{s\ge M} e^{-\sqrt{x}s/2} \\ &\le 1_{-\sqrt{x}<s\le M} e^{-\varepsilon s^2} + 1_{s\ge M} e^{-s/2} \\ &\le e^{-\varepsilon s^2} + e^{-|s|/2} =: F(s) \in L^1(\mathbb{R}, ds). \end{aligned}$$

■

Remark 7.61 (Estimating $q(u)$ by Taylor's Theorem). Another way to estimate $q(u)$ is to use Taylor's theorem with integral remainder. In general if h is C^2 – function on $[0,1]$, then by the fundamental theorem of calculus and integration by parts,

$$\begin{aligned} h(1) - h(0) &= \int_0^1 \dot{h}(t)\,dt = -\int_0^1 \dot{h}(t)\,d(1-t) \\ &= -\dot{h}(t)(1-t)|_0^1 + \int_0^1 \ddot{h}(t)(1-t)\,dt \\ &= \dot{h}(0) + \frac{1}{2}\int_0^1 \ddot{h}(t)\,d\nu(t) \end{aligned} \tag{7.53}$$

where $d\nu(t) := 2(1-t)\,dt$ which is a probability measure on $[0,1]$. Applying this to $h(t) = F(a + t(b-a))$ for a C^2 – function on an interval of points between a and b in $\mathbb{R}$ then implies,

$$F(b) - F(a) = (b-a)\dot{F}(a) + \frac{1}{2}(b-a)^2 \int_0^1 \ddot{F}(a + t(b-a))\,d\nu(t). \tag{7.54}$$

(Similar formulas hold to any order.) Applying this result with $F(x) = x - \ln(1+x)$, $a = 0$, and $b = u \in (-1, \infty)$ gives,

$$u - \ln(1+u) = \frac{1}{2}u^2 \int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t),$$

i.e.

$$q(u) = \frac{1}{2}\int_0^1 \frac{1}{(1+tu)^2}\,d\nu(t).$$

From this expression for $q(u)$ it now easily follows that

$$q(u) \ge \frac{1}{2}\int_0^1 \frac{1}{(1+0)^2}\,d\nu(t) = \frac{1}{2} \text{ if } -1 < u \le 0$$

and

$$q(u) \ge \frac{1}{2}\int_0^1 \frac{1}{(1+u)^2}\,d\nu(t) = \frac{1}{2(1+u)^2}.$$

So an explicit formula for $\varepsilon(M)$ is $\varepsilon(M) = (1+M)^{-2}/2$.

7.6.1 Two applications of Stirling's formula

In this subsection suppose $x \in (0,1)$ and $S_n \overset{d}{=} \text{Binomial}(n, x)$ for all $n \in \mathbb{N}$, i.e.

$$P_x(S_n = k) = \binom{n}{k} x^k (1-x)^{n-k} \text{ for } 0 \le k \le n. \tag{7.55}$$

Recall that $\mathbb{E}S_n = nx$ and $\text{Var}(S_n) = n\sigma^2$ where $\sigma^2 := x(1-x)$. The weak law of large numbers states (Exercise 4.13) that

$$P\left(\left|\frac{S_n}{n} - x\right| \ge \varepsilon\right) \le \frac{1}{n\varepsilon^2}\sigma^2$$

and therefore, $\frac{S_n}{n}$ is concentrating near its mean value, x, for n large, i.e. $S_n \cong nx$ for n large. The next central limit theorem describes the fluctuations of S_n about nx.

Theorem 7.62 (De Moivre-Laplace Central Limit Theorem). *For all $-\infty < a < b < \infty$,*

$$\begin{aligned} \lim_{n\to\infty} P\left(a \le \frac{S_n - nx}{\sigma\sqrt{n}} \le b\right) &= \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}y^2}\,dy \\ &= P(a \le N \le b) \end{aligned}$$

where $N \overset{d}{=} N(0,1)$. Informally, $\frac{S_n - nx}{\sigma\sqrt{n}} \overset{d}{\cong} N$ or equivalently, $S_n \overset{d}{\cong} nx + \sigma\sqrt{n}\cdot N$ which if valid in a neighborhood of nx whose length is order $\sqrt{n}$.

Proof. (We are not going to cover all the technical details in this proof as we will give much more general versions of this theorem later.) Starting with the definition of the Binomial distribution we have,

$$p_n := P\left(a \le \frac{S_n - nx}{\sigma\sqrt{n}} \le b\right) = P\left(S_n \in nx + \sigma\sqrt{n}\,[a,b]\right)$$
$$= \sum_{k \in nx+\sigma\sqrt{n}[a,b]} P(S_n = k)$$
$$= \sum_{k \in nx+\sigma\sqrt{n}[a,b]} \binom{n}{k} x^k (1-x)^{n-k}.$$

Letting $k = nx + \sigma\sqrt{n}y_k$, i.e. $y_k = (k - nx)/\sigma\sqrt{n}$ we see that $\Delta y_k = y_{k+1} - y_k = 1/(\sigma\sqrt{n})$. Therefore we may write p_n as

$$p_n = \sum_{y_k \in [a,b]} \sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \Delta y_k. \tag{7.56}$$

So to finish the proof we need to show, for $k = O(\sqrt{n})$ $(y_k = O(1))$, that

$$\sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \text{ as } n \to \infty \tag{7.57}$$

in which case the sum in Eq. (7.56) may be well approximated by the "Riemann sum;"

$$p_n \sim \sum_{y_k \in [a,b]} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_k^2} \Delta y_k \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy \text{ as } n \to \infty.$$

By Stirling's formula,

$$\sigma\sqrt{n}\binom{n}{k} = \sigma\sqrt{n}\frac{1}{k!}\frac{n!}{(n-k)!} \sim \frac{\sigma\sqrt{n}}{\sqrt{2\pi}} \frac{n^{n+1/2}}{k^{k+1/2}(n-k)^{n-k+1/2}}$$
$$= \frac{\sigma}{\sqrt{2\pi}} \frac{1}{\left(\frac{k}{n}\right)^{k+1/2}\left(1-\frac{k}{n}\right)^{n-k+1/2}}$$
$$= \frac{\sigma}{\sqrt{2\pi}} \frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k+1/2}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k+1/2}}$$
$$\sim \frac{\sigma}{\sqrt{2\pi}} \frac{1}{\sqrt{x(1-x)}} \frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}$$
$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\left(x + \frac{\sigma}{\sqrt{n}}y_k\right)^{k}\left(1 - x - \frac{\sigma}{\sqrt{n}}y_k\right)^{n-k}}.$$

In order to shorten the notation, let $z_k := \frac{\sigma}{\sqrt{n}}y_k = O\left(n^{-1/2}\right)$ so that $k = nx + nz_k = n(x + z_k)$. In this notation we have shown,

$$\sqrt{2\pi}\sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \sim \frac{x^k (1-x)^{n-k}}{(x+z_k)^k (1-x-z_k)^{n-k}}$$
$$= \frac{1}{\left(1 + \frac{1}{x}z_k\right)^k \left(1 - \frac{1}{1-x}z_k\right)^{n-k}}$$
$$= \frac{1}{\left(1 + \frac{1}{x}z_k\right)^{n(x+z_k)} \left(1 - \frac{1}{1-x}z_k\right)^{n(1-x-z_k)}} =: q(n,k). \tag{7.58}$$

Taking logarithms and using Taylor's theorem we learn

$$n(x+z_k)\ln\left(1 + \frac{1}{x}z_k\right)$$
$$= n(x+z_k)\left(\frac{1}{x}z_k - \frac{1}{2x^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= nz_k + \frac{n}{2x}z_k^2 + O\left(n^{-3/2}\right) \text{ and}$$
$$n(1-x-z_k)\ln\left(1 - \frac{1}{1-x}z_k\right)$$
$$= n(1-x-z_k)\left(-\frac{1}{1-x}z_k - \frac{1}{2(1-x)^2}z_k^2 + O\left(n^{-3/2}\right)\right)$$
$$= -nz_k + \frac{n}{2(1-x)}z_k^2 + O\left(n^{-3/2}\right).$$

and then adding these expressions shows,

$$-\ln q(n,k) = \frac{n}{2}z_k^2\left(\frac{1}{x} + \frac{1}{1-x}\right) + O\left(n^{-3/2}\right)$$
$$= \frac{n}{2\sigma^2}z_k^2 + O\left(n^{-3/2}\right) = \frac{1}{2}y_k^2 + O\left(n^{-3/2}\right).$$

Combining this with Eq. (7.58) shows,

$$\sigma\sqrt{n}\binom{n}{k} x^k (1-x)^{n-k} \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_k^2 + O\left(n^{-3/2}\right)\right)$$

which gives the desired estimate in Eq. (7.57). ■

The previous central limit theorem has shown that

$$\frac{S_n}{n} \overset{d}{\cong} x + \frac{\sigma}{\sqrt{n}}N$$

which implies the major fluctuations of S_n/n occur within intervals about x of length $O\left(\frac{1}{\sqrt{n}}\right)$. The next result aims to understand the rare events where S_n/n makes a "large" deviation from its mean value, x – in this case a large deviation is something of size $O(1)$ as $n \to \infty$.

Theorem 7.63 (Binomial Large Deviation Bounds). *Let us continue to use the notation in Theorem 7.62. Then for all $y \in (0, x)$,*

$$\lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) = y \ln \frac{x}{y} + (1-y) \ln \frac{1-x}{1-y}.$$

Roughly speaking,

$$P_x\left(\frac{S_n}{n} \le y\right) \approx e^{-nI_x(y)}$$

where $I_x(y)$ is the "rate function,"

$$I_x(y) := y \ln \frac{y}{x} + (1-y) \ln \frac{1-y}{1-x},$$

see Figure 7.3 for the graph of $I_{1/2}$.

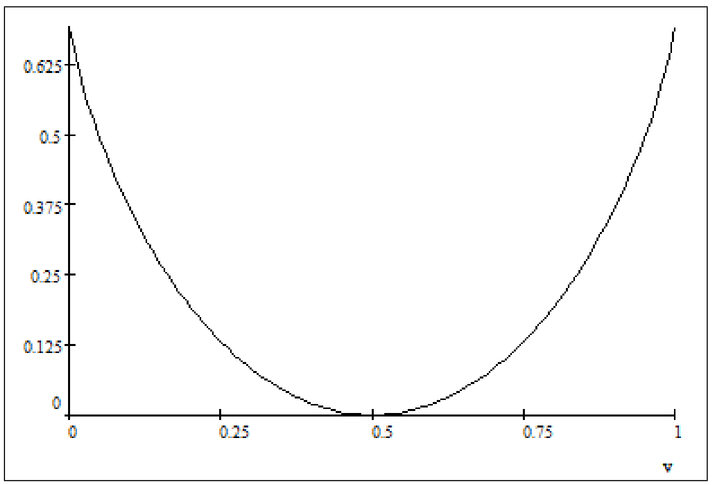


Fig. 7.3. A plot of the rate function, $I_{1/2}$.

Proof. By definition of the binomial distribution,

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

If $a_k \ge 0$, then we have the following crude estimates on $\sum_{k=0}^{m-1} a_k$,

$$\max_{k<m} a_k \le \sum_{k=0}^{m-1} a_k \le m \cdot \max_{k<m} a_k. \tag{7.59}$$

In order to apply this with $a_k = \binom{n}{k} x^k (1-x)^{n-k}$ and $m = [ny]$, we need to find the maximum of the a_k for $0 \le k \le ny$. This is easy to do since a_k is increasing for $0 \le k \le ny$ as we now show. Consider,

$$\begin{aligned}\frac{a_{k+1}}{a_k} &= \frac{\binom{n}{k+1} x^{k+1} (1-x)^{n-k-1}}{\binom{n}{k} x^k (1-x)^{n-k}} \\ &= \frac{k!\,(n-k)! \cdot x}{(k+1)! \cdot (n-k-1)! \cdot (1-x)} \\ &= \frac{(n-k) \cdot x}{(k+1) \cdot (1-x)}.\end{aligned}$$

Therefore, where the latter expression is greater than or equal to 1 iff

$$\begin{aligned}\frac{a_{k+1}}{a_k} \ge 1 &\iff (n-k) \cdot x \ge (k+1) \cdot (1-x) \\ &\iff nx \ge k+1-x \iff k < (n-1)x - 1.\end{aligned}$$

Thus for $k < (n-1)x - 1$ we may conclude that $\binom{n}{k} x^k (1-x)^{n-k}$ is increasing in k.

Thus the crude bound in Eq. (7.59) implies,

$$\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]} \le P_x\left(\frac{S_n}{n} \le y\right) \le [ny] \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}$$

or equivalently,

$$\begin{aligned}&\frac{1}{n} \ln \left[\binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right] \\ &\qquad \le \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) \\ &\qquad\qquad \le \frac{1}{n} \ln \left[(ny) \binom{n}{[ny]} x^{[ny]} (1-x)^{n-[ny]}\right].\end{aligned}$$

By Stirling's formula, for k such that k and $n-k$ is large we have,

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi}} \frac{n^{n+1/2}}{k^{k+1/2} \cdot (n-k)^{n-k+1/2}} = \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{1}{\left(\frac{k}{n}\right)^{k+1/2} \cdot \left(1-\frac{k}{n}\right)^{n-k+1/2}}$$

and therefore,

$$\frac{1}{n} \ln \binom{n}{k} \sim -\frac{k}{n} \ln\left(\frac{k}{n}\right) - \left(1-\frac{k}{n}\right) \ln\left(1-\frac{k}{n}\right).$$

So taking $k = [ny]$, we learn that

$$\lim_{n\to\infty} \frac{1}{n} \ln \binom{n}{[ny]} = -y \ln y - (1-y) \ln(1-y)$$

and therefore,

$$\lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) = -y\ln y - (1-y)\ln(1-y) + y\ln x + (1-y)\ln(1-x)$$
$$= y \ln \frac{x}{y} + (1-y) \ln\left(\frac{1-x}{1-y}\right).$$

■

As a consistency check it is worth noting, by Jensen's inequality described below, that

$$-I_x(y) = y \ln \frac{x}{y} + (1-y) \ln\left(\frac{1-x}{1-y}\right) \le \ln\left(y\frac{x}{y} + (1-y)\frac{1-x}{1-y}\right) = \ln(1) = 0.$$

This must be the case since

$$-I_x(y) = \lim_{n\to\infty} \frac{1}{n} \ln P_x\left(\frac{S_n}{n} \le y\right) \le \lim_{n\to\infty} \frac{1}{n} \ln 1 = 0.$$

7.7 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

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{7.60}$$

of $[a,b]$ let

$$\text{mesh}(\pi) := \max\{|t_j - t_{j-1}| : j = 1, \ldots, n\},$$

$$M_j = \sup\{f(x) : t_j \le x \le t_{j-1}\}, \quad m_j = \inf\{f(x) : t_j \le x \le t_{j-1}\}$$

$$G_\pi = f(a)1_{\{a\}} + \sum_1^n M_j 1_{(t_{j-1},t_j]}, \quad g_\pi = f(a)1_{\{a\}} + \sum_1^n m_j 1_{(t_{j-1},t_j]} \text{ and}$$

$$S_\pi f = \sum M_j(t_j - t_{j-1}) \text{ and } s_\pi f = \sum m_j(t_j - t_{j-1}).$$

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 7.64. *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 7.29.

Lemma 7.65. *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 $\text{mesh}(\pi_k) \downarrow 0$ *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

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{7.61}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} g dm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{7.62}$$

and

$$\int_{[a,b]} G dm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{7.63}$$

Notation 7.66 *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 7.67. *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 $\mathrm{mesh}(\pi_k) \downarrow 0$ and G and g are defined as in Eq. (7.61), then*
$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall\, x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{7.64}$$
(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
$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x)\ \forall\ x \notin \pi. \tag{7.65}$$
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. (7.65) 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. (7.64) 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 7.68. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{7.66}$$

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[2]*, 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*

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{7.67}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 7.65 and let G and g be defined as in Lemma 7.67. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (7.66) is a consequence of Eqs. (7.62) and (7.63). From Eq. (7.66), 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 $E = \{x : H(x) \ne h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (7.64), the remaining assertions including Eq. (7.67) are now consequences of Lemma 7.71. ∎

Notation 7.69 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

[2] f need not be Borel measurable.

7.8 Measurability on Complete Measure Spaces*

In this subsection we will discuss a couple of measurability results concerning completions of measure spaces.

Proposition 7.70. *Suppose that $(X,\mathcal{B},\mu)$ is a complete measure space*[3] *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,

$$(1_E g)^{-1}(A)=\begin{cases}E^c\cup(1_E g)^{-1}(A\setminus\{0\}) & \text{if } 0\in A\\ (1_E g)^{-1}(A) & \text{if } 0\notin A\end{cases} \tag{7.68}$$

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. (7.68) shows that $1_E g$ is measurable. 2. Let $E=\{x:\lim\limits_{n\to\infty}f_n(x)\neq f(x)\}$ 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 7.71. *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 $g:X\to\mathbb{R}$ 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*

$$\int_X fd\bar\mu=\int_X gd\mu.$$

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 7.70 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 $f\ge 0$. Choose $(\bar{\mathcal{M}},\mathcal{B})$ – measurable simple function $\varphi_n\ge 0$ 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 $\tilde\varphi_n\ge 0$ such that $E_n:=\{\varphi_n\neq\tilde\varphi_n\}$ has zero $\bar\mu$ – measure. Since $\bar\mu(\cup_n E_n)\le\sum_n\bar\mu(E_n)$, 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., $\int_X fd\bar\mu=\int_X gd\bar\mu$ so to prove Eq. (7.69) it suffices to prove

$$\int_X gd\bar\mu=\int_X gd\mu. \tag{7.69}$$

Because $\bar\mu=\mu$ on $\mathcal{M}$, Eq. (7.69) 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 $(\operatorname{Im}g)_\pm$. ■

[3] 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.

7.9 More Exercises

Exercise 7.19. 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 7.20 (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:
$$\rho(A,C)\le\rho(A,B)+\rho(B,C) \text{ for all } A,B,C\in\mathcal{M}.$$
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 7.21. 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 7.22. 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 7.23. Let $(X, \mathcal{M}, \mu)$ be a measure space and $\{A_n\}_{n=1}^\infty \subset \mathcal{M}$, show

$$\mu(\{A_n \text{ a.a.}\}) \le \liminf_{n\to\infty} \mu(A_n)$$

and if $\mu(\cup_{m\ge n} A_m) < \infty$ for some n, then

$$\mu(\{A_n \text{ i.o.}\}) \ge \limsup_{n\to\infty} \mu(A_n).$$

Exercise 7.24 (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 7.25. 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 7.26. 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 7.27. 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 7.28. 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 7.29. Prove Lemma 7.65.

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.

8.1 Multiplicative System Theorems

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 [1, p. 14]. The style of proof given here may be found in Janson [5, 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. 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 $\sigma(\mathbb{M}) \subset \mathcal{B}$.

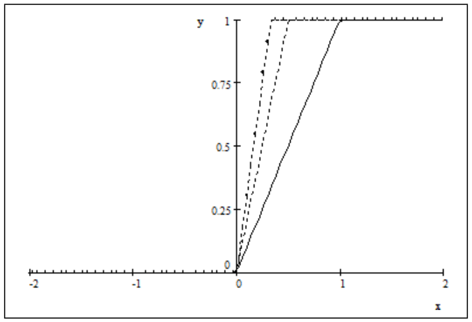


Fig. 8.1. Plots of φ_1, φ_2 and φ_3.

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 $\lambda -$ 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 $\pi -$ 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. ∎

Corollary 8.3. *Suppose $\mathbb{H}$ is a subspace of bounded real valued 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 8.2. ∎

Example 8.4. 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 8.2 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 $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

Here is a complex version of Theorem 8.2.

Theorem 8.5 (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 $\mathbb{M}_0 = \text{span}_\mathbb{C}(\mathbb{M} \cup \{1\})$ 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 8.2, $\mathbb{H}^\mathbb{R}$ 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 $\text{Re} f = \frac{1}{2}(f + \bar{f})$ and $\text{Im} f = \frac{1}{2i}(f - \bar{f})$ 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 $f = \text{Re} f + i \,\text{Im} f \in \mathbb{H}$ since $\text{Re} f$ and $\text{Im} f$ are in $\mathbb{H}^\mathbb{R}$. ∎

Lemma 8.6. *Suppose that $-\infty < a < b < \infty$ and let Trig$(\mathbb{R}) \subset C(\mathbb{R}, \mathbb{C})$ be the complex linear span of $\{x \to e^{i\lambda x} : \lambda \in \mathbb{R}\}$. Then there exists $f_n \in C_c(\mathbb{R}, [0,1])$ and $g_n \in$ Trig$(\mathbb{R})$ such that $\lim_{n\to\infty} f_n(x) = 1_{(a,b]}(x) = \lim_{n\to\infty} g_n(x)$ for all $x \in \mathbb{R}$.*

Proof. The assertion involving $f_n \in C_c(\mathbb{R}, [0,1])$ was the content of one of your homework assignments. For the assertion involving $g_n \in$Trig$(\mathbb{R})$, it will suffice to show that any $f \in C_c(\mathbb{R})$ may be written as $f(x) = \lim_{n\to\infty} g_n(x)$

for some $\{g_n\} \subset \mathrm{Trig}(\mathbb{R})$ where the limit is uniform for x in compact subsets of $\mathbb{R}$.

So suppose that $f \in C_c(\mathbb{R})$ and $L > 0$ such that $f(x) = 0$ if $|x| \geq L/4$. Then

$$f_L(x) := \sum_{n=-\infty}^{\infty} f(x + nL)$$

is a continuous L – periodic function on $\mathbb{R}$, see Figure 8.2. If $\varepsilon > 0$ is given, we

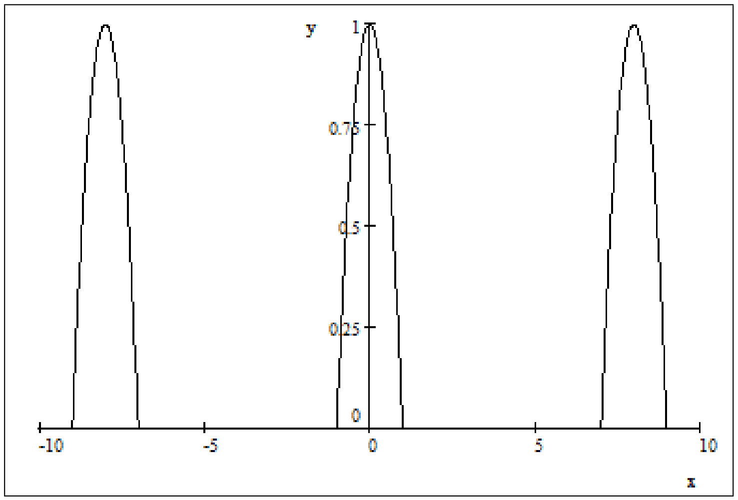


Fig. 8.2. This is plot of $f_8(x)$ where $f(x) = \left(1 - x^2\right) 1_{|x| \leq 1}$. The center hump by itself would be the plot of $f(x)$.

may apply Theorem 4.42 to find $\Lambda \subset\subset \mathbb{Z}$ such that

$$\left| f_L\left(\frac{L}{2\pi} x\right) - \sum_{\alpha \in \Lambda} a_\lambda e^{i\alpha x} \right| \leq \varepsilon \text{ for all } x \in \mathbb{R},$$

wherein we have use the fact that $x \to f_L\left(\frac{L}{2\pi} x\right)$ is a 2π – periodic function of x. Equivalently we have,

$$\max_x \left| f_L(x) - \sum_{\alpha \in \Lambda} a_\lambda e^{i \frac{2\pi\alpha}{L} x} \right| \leq \varepsilon.$$

In particular it follows that $f_L(x)$ is a uniform limit of functions from $\mathrm{Trig}(\mathbb{R})$. Since $\lim_{L\to\infty} f_L(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$, it is easy to conclude there exists $g_n \in \mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n(x) = f(x)$ uniformly on compact subsets of $\mathbb{R}$. ■

Corollary 8.7. *Each of the following σ – algebras on $\mathbb{R}^d$ are equal to $\mathcal{B}_{\mathbb{R}^d}$;*

1. $\mathcal{M}_1 := \sigma\left(\cup_{i=1}^n \{x \to f(x_i) : f \in C_c(\mathbb{R})\}\right)$,
2. $\mathcal{M}_2 := \sigma\left(x \to f_1(x_1) \dots f_d(x_d) : f_i \in C_c(\mathbb{R})\right)$
3. $\mathcal{M}_3 = \sigma\left(C_c\left(\mathbb{R}^d\right)\right)$, *and*
4. $\mathcal{M}_4 := \sigma\left(\left\{x \to e^{i\lambda \cdot x} : \lambda \in \mathbb{R}^d\right\}\right)$.

Proof. As the functions defining each $\mathcal{M}_i$ are continuous and hence Borel measurable, it follows that $\mathcal{M}_i \subset \mathcal{B}_{\mathbb{R}^d}$ for each i. So to finish the proof it suffices to show $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_i$ for each i.

$\mathcal{M}_1$ case. Let $a, b \in \mathbb{R}$ with $-\infty < a < b < \infty$. By Lemma 8.6, there exists $f_n \in C_c(\mathbb{R})$ such that $\lim_{n\to\infty} f_n = 1_{(a,b]}$. Therefore it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_1$ – measurable for each i. Moreover if $-\infty < a_i < b_i < \infty$ for each i, then we may conclude that

$$x \to \prod_{i=1}^d 1_{(a_i,b_i]}(x_i) = 1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x)$$

is $\mathcal{M}_1$ – measurable as well and hence $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_1$. As such sets generate $\mathcal{B}_{\mathbb{R}^d}$ we may conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_1$.

and therefore $\mathcal{M}_1 = \mathcal{B}_{\mathbb{R}^d}$.

$\mathcal{M}_2$ case. As above, we may find $f_{i,n} \to 1_{(a_i,b_i]}$ as $n \to \infty$ for each $1 \leq i \leq d$ and therefore,

$$1_{(a_1,b_1]\times\dots\times(a_d,b_d]}(x) = \lim_{n\to\infty} f_{1,n}(x_1) \dots f_{d,n}(x_d) \text{ for all } x \in \mathbb{R}^d.$$

This shows that $1_{(a_1,b_1]\times\dots\times(a_d,b_d]}$ is $\mathcal{M}_2$ – measurable and therefore $(a_1, b_1] \times \dots \times (a_d, b_d] \in \mathcal{M}_2$.

$\mathcal{M}_3$ case. This is easy since $\mathcal{B}_{\mathbb{R}^d} = \mathcal{M}_2 \subset \mathcal{M}_3$.

$\mathcal{M}_4$ case. By Lemma 8.6 here exists $g_n \in \mathrm{Trig}(\mathbb{R})$ such that $\lim_{n\to\infty} g_n = 1_{(a,b]}$. Since $x \to g_n(x_i)$ is in the span $\left\{x \to e^{i\lambda \cdot x} : \lambda \in \mathbb{R}^d\right\}$ for each n, it follows that $x \to 1_{(a,b]}(x_i)$ is $\mathcal{M}_4$ – measurable for all $-\infty < a < b < \infty$. Therefore, just as in the proof of case 1., we may now conclude that $\mathcal{B}_{\mathbb{R}^d} \subset \mathcal{M}_4$. ■

Corollary 8.8. *Suppose that $\mathbb{H}$ is a subspace of complex valued functions on $\mathbb{R}^d$ which is closed under complex conjugation and bounded convergence. If $\mathbb{H}$ contains any one of the following collection of functions;*

1. $\mathbb{M} := \{x \to f_1(x_1) \dots f_d(x_d) : f_i \in C_c(\mathbb{R})\}$
2. $\mathbb{M} := C_c\left(\mathbb{R}^d\right)$, *or*
3. $\mathbb{M} := \left\{x \to e^{i\lambda \cdot x} : \lambda \in \mathbb{R}^d\right\}$

then $\mathbb{H}$ contains all bounded complex Borel measurable functions on $\mathbb{R}^d$.

Proof. Observe that if $f \in C_c(\mathbb{R})$ such that $f(x) = 1$ in a neighborhood of 0, then $f_n(x) := f(x/n) \to 1$ as $n \to \infty$. Therefore in cases 1. and 2., $\mathbb{H}$ contains the constant function, 1, since

$$1 = \lim_{n\to\infty} f_n(x_1)\dots f_n(x_d).$$

In case 3, $1 \in \mathbb{M} \subset \mathbb{H}$ as well. The result now follows from Theorem 8.5 and Corollary 8.7. ∎

Proposition 8.9 (Change of Variables Formula). *Suppose that $-\infty < a < b < \infty$ and $u : [a,b] \to \mathbb{R}$ is a continuously differentiable function. Let $[c,d] = u([a,b])$ where $c = \min u([a,b])$ and $d = \max u([a,b])$. (By the intermediate value theorem $u([a,b])$ is an interval.) Then for all bounded measurable functions, $f : [c,d] \to \mathbb{R}$ we have*

$$\int_{u(a)}^{u(b)} f(x)\,dx = \int_a^b f(u(t))\,\dot{u}(t)\,dt. \tag{8.2}$$

Moreover, Eq. (8.2) is also valid if $f : [c,d] \to \mathbb{R}$ is measurable and

$$\int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty. \tag{8.3}$$

Proof. Let $\mathbb{H}$ denote the space of bounded measurable functions such that Eq. (8.2) holds. It is easily checked that $\mathbb{H}$ is a linear space closed under bounded convergence. Next we show that $\mathbb{M} = C([c,d],\mathbb{R}) \subset \mathbb{H}$ which coupled with Corollary 8.8 will show that $\mathbb{H}$ contains all bounded measurable functions from $[c,d]$ to $\mathbb{R}$.

If $f : [c,d] \to \mathbb{R}$ is a continuous function and let F be an anti-derivative of f. Then by the fundamental theorem of calculus,

$$\begin{aligned}\int_a^b f(u(t))\,\dot{u}(t)\,dt &= \int_a^b F'(u(t))\,\dot{u}(t)\,dt \\ &= \int_a^b \frac{d}{dt}F(u(t))\,dt = F(u(t))|_a^b \\ &= F(u(b)) - F(u(a)) = \int_{u(a)}^{u(b)} F'(x)\,dx = \int_{u(a)}^{u(b)} f(x)\,dx.\end{aligned}$$

Thus $\mathbb{M} \subset \mathbb{H}$ and the first assertion of the proposition is proved.

Now suppose that $f : [c,d] \to \mathbb{R}$ is measurable and Eq. (8.3) holds. For $M < \infty$, let $f_M(x) = f(x)\cdot 1_{|f(x)|\le M}$ – a bounded measurable function. Therefore applying Eq. (8.2) with f replaced by $|f_M|$ shows,

$$\left|\int_{u(a)}^{u(b)} |f_M(x)|\,dx\right| = \left|\int_a^b |f_M(u(t))|\,\dot{u}(t)\,dt\right| \le \int_a^b |f_M(u(t))|\,|\dot{u}(t)|\,dt.$$

Using the MCT, we may let $M \uparrow \infty$ in the previous inequality to learn

$$\left|\int_{u(a)}^{u(b)} |f(x)|\,dx\right| \le \int_a^b |f(u(t))|\,|\dot{u}(t)|\,dt < \infty.$$

Now apply Eq. (8.2) with f replaced by f_M to learn

$$\int_{u(a)}^{u(b)} f_M(x)\,dx = \int_a^b f_M(u(t))\,\dot{u}(t)\,dt.$$

Using the DCT we may now let $M \to \infty$ in this equation to show that Eq. (8.2) remains valid. ∎

Exercise 8.1. Suppose that $u : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that $\dot{u}(t) \ge 0$ for all t and $\lim_{t\to\pm\infty} u(t) = \pm\infty$. Show that

$$\int_{\mathbb{R}} f(x)\,dx = \int_{\mathbb{R}} f(u(t))\,\dot{u}(t)\,dt \tag{8.4}$$

for all measurable functions $f : \mathbb{R} \to [0,\infty]$. In particular applying this result to $u(t) = at + b$ where $a > 0$ implies,

$$\int_{\mathbb{R}} f(x)\,dx = a\int_{\mathbb{R}} f(at+b)\,dt.$$

Definition 8.10. *The **Fourier transform** or **characteristic function** of a finite measure, μ, on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, is the function, $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^d} e^{i\lambda\cdot x}\,d\mu(x) \text{ for all } \lambda \in \mathbb{R}^d$$

Corollary 8.11. *Suppose that μ and ν are two probability measures on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Then any one of the next three conditions implies that $\mu = \nu$;*

1. $\int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\nu(x) = \int_{\mathbb{R}^d} f_1(x_1)\dots f_d(x_d)\,d\mu(x)$ *for all* $f_i \in C_c(\mathbb{R})$.
2. $\int_{\mathbb{R}^d} f(x)\,d\nu(x) = \int_{\mathbb{R}^d} f(x)\,d\mu(x)$ *for all* $f \in C_c(\mathbb{R}^d)$.
3. $\hat{\nu} = \hat{\mu}$.

Item 3. asserts that the Fourier transform is injective.

Proof. Let $\mathbb{H}$ be the collection of bounded complex measurable functions from $\mathbb{R}^d$ to $\mathbb{C}$ such that

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\nu. \tag{8.5}$$

It is easily seen that $\mathbb{H}$ is a linear space closed under complex conjugation and bounded convergence (by the DCT). Since $\mathbb{H}$ contains one of the multiplicative systems appearing in Corollary 8.8, it contains all bounded Borel measurable functions form $\mathbb{R}^d \to \mathbb{C}$. Thus we may take $f = 1_A$ with $A \in \mathcal{B}_{\mathbb{R}^d}$ in Eq. (8.5) to learn, $\mu(A) = \nu(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$. ■

In many cases we can replace the condition in item 3. of Corollary 8.11 by;

$$\int_{\mathbb{R}^d} e^{\lambda \cdot x} d\mu(x) = \int_{\mathbb{R}^d} e^{\lambda \cdot x} d\nu(x) \text{ for all } \lambda \in U, \tag{8.6}$$

where U is a neighborhood of $0 \in \mathbb{R}^d$. In order to do this, one must assume at least assume that the integrals involved are finite for all $\lambda \in U$. The idea is to show that Condition 8.6 implies $\hat{\nu} = \hat{\mu}$. You are asked to carry out this argument in Exercise 8.2 making use of the following lemma.

Lemma 8.12 (Analytic Continuation). *Let $\varepsilon > 0$ and $S_\varepsilon := \{x + iy \in \mathbb{C} : |x| < \varepsilon\}$ be an ε strip in $\mathbb{C}$ about the imaginary axis. Suppose that $h : S_\varepsilon \to \mathbb{C}$ is a function such that for each $b \in \mathbb{R}$, there exists $\{c_n(b)\}_{n=0}^\infty \subset \mathbb{C}$ such that*

$$h(z + ib) = \sum_{n=0}^\infty c_n(b) z^n \text{ for all } |z| < \varepsilon. \tag{8.7}$$

If $c_n(0) = 0$ for all $n \in \mathbb{N}_0$, then $h \equiv 0$.

Proof. It suffices to prove the following assertion; if for some $b \in \mathbb{R}$ we know that $c_n(b) = 0$ for all n, then $c_n(y) = 0$ for all n and $y \in (b - \varepsilon, b + \varepsilon)$. We now prove this assertion.

Let us assume that $b \in \mathbb{R}$ and $c_n(b) = 0$ for all $n \in \mathbb{N}_0$. It then follows from Eq. (8.7) that $h(z + ib) = 0$ for all $|z| < \varepsilon$. Thus if $|y - b| < \varepsilon$, we may conclude that $h(x + iy) = 0$ for x in a (possibly very small) neighborhood $(-\delta, \delta)$ of 0. Since

$$\sum_{n=0}^\infty c_n(y) x^n = h(x + iy) = 0 \text{ for all } |x| < \delta,$$

it follows that

$$0 = \frac{1}{n!} \frac{d^n}{dx^n} h(x + iy)|_{x=0} = c_n(y)$$

and the proof is complete. ■

8.2 Exercises

Exercise 8.2. Suppose $\varepsilon > 0$ and X and Y are two random variables such that $\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{tY}\right] < \infty$ for all $|t| \le \varepsilon$. Show;

1. $\mathbb{E}\left[e^{\varepsilon|X|}\right]$ and $\mathbb{E}\left[e^{\varepsilon|Y|}\right]$ are finite.
2. $\mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{itY}\right]$ for all $t \in \mathbb{R}$. **Hint:** Consider $h(z) := \mathbb{E}\left[e^{zX}\right] - \mathbb{E}\left[e^{zY}\right]$ for $z \in S_\varepsilon$. Now show for $|z| \le \varepsilon$ and $b \in \mathbb{R}$, that
$$h(z + ib) = \mathbb{E}\left[e^{ibX} e^{zX}\right] - \mathbb{E}\left[e^{ibY} e^{zY}\right] = \sum_{n=0}^\infty c_n(b) z^n \tag{8.8}$$
where
$$c_n(b) := \frac{1}{n!}\left(\mathbb{E}\left[e^{ibX} X^n\right] - \mathbb{E}\left[e^{ibY} Y^n\right]\right). \tag{8.9}$$
3. Conclude from item 2. that $X \overset{d}{=} Y$, i.e. that $\operatorname{Law}_P(X) = \operatorname{Law}_P(Y)$.

Exercise 8.3. 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 $h(x, y) = 1_{\{x=y\}}$.

Exercise 8.4 (Density of $\mathcal{A}$ – simple functions). Let $(\Omega, \mathcal{B}, P)$ be a probability space and assume that $\mathcal{A}$ is a sub-algebra of $\mathcal{B}$ such that $\mathcal{B} = \sigma(\mathcal{A})$. Let $\mathbb{H}$ denote the bounded measurable functions $f : \Omega \to \mathbb{R}$ such that for every $\varepsilon > 0$ there exists an an $\mathcal{A}$ – simple function, $\varphi : \Omega \to \mathbb{R}$ such that $\mathbb{E}|f - \varphi| < \varepsilon$. Show $\mathbb{H}$ consists of all bounded measurable functions, $f : \Omega \to \mathbb{R}$. **Hint:** let $\mathbb{M}$ denote the collection of $\mathcal{A}$ – simple functions.

Corollary 8.13. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $\{X_n\}_{n=1}^\infty$ is a collection of random variables on Ω, and $\mathcal{B}_\infty := \sigma(X_1, X_2, X_3, \dots)$. Then for all $\varepsilon > 0$ and all bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, there exists an $n \in \mathbb{N}$ and a bounded $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}|f - G(X_1, \dots, X_n)| < \varepsilon$. Moreover we may assume that $\sup_{x \in \mathbb{R}^n} |G(x)| \le M := \sup_{\omega \in \Omega} |f(\omega)|$.*

Proof. Apply Exercise 8.4 with $\mathcal{A} := \cup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$ in order to find an $\mathcal{A}$ – measurable simple function, φ, such that $\mathbb{E}|f - \varphi| < \varepsilon$. By the definition of $\mathcal{A}$ we know that φ is $\sigma(X_1, \dots, X_n)$ – measurable for some $n \in \mathbb{N}$. It now follows by the factorization Lemma 6.40 that $\varphi = G(X_1, \dots, X_n)$ for some $\mathcal{B}_{\mathbb{R}^n}$ – measurable function $G : \mathbb{R}^n \to \mathbb{R}$. If necessary, replace G by $[G \wedge M] \vee (-M)$ in order to insure $\sup_{x \in \mathbb{R}^n} |G(x)| \le M$. ■

Exercise 8.5 (Density of $\mathcal{A}$ in $\mathcal{B} = \sigma(\mathcal{A})$). Keeping the same notation as in Exercise 8.4 but now take $f = 1_B$ for some $B \in \mathcal{B}$ and given $\varepsilon > 0$, write $\varphi = \sum_{i=0}^{n} \lambda_i 1_{A_i}$ where $\lambda_0 = 0$, $\{\lambda_i\}_{i=1}^{n}$ is an enumeration of $\varphi(\Omega) \setminus \{0\}$, and $A_i := \{\varphi = \lambda_i\}$. Show; 1.

$$\mathbb{E}|1_B - \varphi| = P(A_0 \cap B) + \sum_{i=1}^{n} [|1 - \lambda_i| P(B \cap A_i) + |\lambda_i| P(A_i \setminus B)] \quad (8.10)$$

$$\ge P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\}. \quad (8.11)$$

2. Now let $\psi = \sum_{i=0}^{n} \alpha_i 1_{A_i}$ with

$$\alpha_i = \begin{cases} 1 \text{ if } P(A_i \setminus B) \le P(B \cap A_i) \\ 0 \text{ if } P(A_i \setminus B) > P(B \cap A_i) \end{cases}.$$

Then show that

$$\mathbb{E}|1_B - \psi| = P(A_0 \cap B) + \sum_{i=1}^{n} \min\{P(B \cap A_i), P(A_i \setminus B)\} \le \mathbb{E}|1_B - \varphi|.$$

Observe that $\psi = 1_D$ where $D = \cup_{i:\alpha_{i=1}} A_i \in \mathcal{A}$ and so you have shown; for every $\varepsilon > 0$ there exists a $D \in \mathcal{A}$ such that

$$P(B \Delta D) = \mathbb{E}|1_B - 1_D| < \varepsilon.$$

8.3 A Strengthening of the Multiplicative System Theorem*

Notation 8.14 *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 [13, p. 365].

Proposition 8.15. **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 8.16. **Suppose that $\mathbb{H}$ is a vector subspace of bounded real valued functions on Ω 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 8.15 reduces this theorem to Theorem 8.2. ■

8.4 The Bounded Approximation Theorem*

This section should be skipped until needed (if ever!).

Notation 8.17 *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 8.18 (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 $\mathbb{M}_\downarrow \ni f_n \le f \le f \in \mathbb{M}_\uparrow$ 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

[1] Bruce: rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

$$\mu(\lambda h - \lambda f) = \lambda \mu(h - f) < \lambda \varepsilon.$$
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 8.16, 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,
$$\begin{aligned}\mu\left(v - u^N\right) &= \sum_{n=1}^{N} \mu(v_n - u_n) + \sum_{n=N+1}^{\infty} \mu(v_n) \le \sum_{n=1}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \mu(v_n) \\ &\le \varepsilon + \sum_{n=N+1}^{\infty} \mu(v_n).\end{aligned}$$
However, since
$$\begin{aligned}\sum_{n=1}^{\infty} \mu(v_n) &\le \sum_{n=1}^{\infty} \mu\left(\delta_n + \varepsilon 2^{-n}\right) = \sum_{n=1}^{\infty} \mu(\delta_n) + \varepsilon \mu(\Omega) \\ &= \sum_{n=1}^{\infty} \mu(g - g_1) + \varepsilon \mu(\Omega) < \infty,\end{aligned}$$
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 8.3 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

$$I^2 = \left[\int_{\mathbb{R}} e^{-x^2/2}dm(x)\right]^2 = \int_{\mathbb{R}} e^{-y^2/2}\left[\int_{\mathbb{R}} e^{-x^2/2}dm(x)\right]dm(y)$$
$$= \int_{\mathbb{R}^2} e^{-\left(x^2+y^2\right)/2}dm^2(x,y)$$

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,

$$I^2 = \lim_{R\to\infty}\int_{D_R} e^{-\left(x^2+y^2\right)/2}dm^2(x,y)$$

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

$$\int_{D_R} e^{-\left(x^2+y^2\right)/2}d\pi(x,y) = \int_{(0,R)\times(0,2\pi)} e^{-r^2/2}r\,dr\,d\theta$$
$$= 2\pi\int_0^R e^{-r^2/2}r\,dr = 2\pi\left(1-e^{-R^2/2}\right).$$

[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

Notation 9.8 *If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is any measurable function, let*

$$\bar{\int}_X f d\mu := \begin{cases} \int_X f d\mu & \text{if} \quad \int_X |f|\, d\mu < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 9.9 (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, $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$, $\bar{\int}_X f(x, \cdot)d\mu(x) \in L^1(\nu)$ and Eqs. (9.8) and (9.9) are still valid after putting a bar over the integral symbols.

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

$$\begin{aligned} \bar{\int}_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) \\ &= \int_Y 1_{E^c}(x) \left[f_+(x, y) - f_-(x, y)\right] d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x, y)\, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x, y)\, d\nu(y). \end{aligned} \tag{9.14}$$

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 $x \to \bar{\int}_Y f(x, y)\, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

$$\int_X \left|\bar{\int}_Y f(x, y)\, d\nu(y)\right| d\mu(x) \le \int_X \left[\int_Y |f(x, y)|\, d\nu(y)\right] d\mu(x) < \infty,$$

which shows $\bar{\int}_Y f(\cdot, y)dv(y) \in L^1(\mu)$. Integrating Eq. (9.14) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned} &\int_X \left[\bar{\int}_Y f(x, y)\, d\nu(y)\right] d\mu(x) \\ &= \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_+(x, y) - \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x) f_-(x, y) \\ &= \int_Y d\nu(y) \int_X d\mu(x) f_+(x, y) - \int_Y d\nu(y) \int_X d\mu(x) f_-(x, y) \\ &= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-)\, d\pi = \int_{X\times Y} f d\pi \end{aligned} \tag{9.15}$$

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 $\mu(E) = 0$ and

$$\begin{aligned} \bar{\int}_Y f(x, y)\, d\nu(y) &= \int_Y 1_{E^c}(x) f(x, y)\, d\nu(y) = \int_Y 1_{E^c}(x) \left[u(x, y) + iv(x, y)\right] d\nu(y) \\ &= \int_Y 1_{E^c}(x) u(x, y)\, d\nu(y) + i \int_Y 1_{E^c}(x) v(x, y)\, d\nu(y). \end{aligned}$$

The last line is a measurable in x as we have just proved. Similarly one shows $\bar{\int}_Y f(\cdot, y)\, d\nu(y) \in L^1(\mu)$ 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 generalize to products of any finite number of σ – finite measure spaces.

Theorem 9.10. *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*

$$\pi(A_1 \times \cdots \times A_n) = \mu_1(A_1) \ldots \mu_n(A_n) \text{ for all } A_i \in \mathcal{M}_i. \tag{9.16}$$

(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*

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1, \ldots, x_n) \tag{9.17}$$

where σ *is any permutation of* $\{1, 2, \ldots, n\}$. *In particular* $f \in L^1(\pi)$, *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. *Furthermore, if* $f \in L^1(\pi)$, *then*

$$\int_X f d\pi = \bar{\int}_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \bar{\int}_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\ f(x_1, \ldots, x_n) \tag{9.18}$$

for all permutations σ.

Proof. (* I would consider skipping this tedious proof.) The proof will be by induction on n with the case $n = 2$ being covered in Theorems 9.6 and 9.9. So let $n \geq 3$ and assume the theorem is valid for $n - 1$ factors or less. To simplify notation, for $1 \leq i \leq n$, let $X^i = \prod_{j \neq i} X_j$, $\mathcal{M}^i := \otimes_{j \neq i} \mathcal{M}_i$, and $\mu^i := \otimes_{j \neq i} \mu_j$ be the product measure on $(X^i, \mathcal{M}^i)$ which is assumed to exist by the induction hypothesis. Also let $\mathcal{M} := \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ and for $x = (x_1, \ldots, x_i, \ldots, x_n) \in X$ let

$$x^i := (x_1, \ldots, \hat{x}_i, \ldots, x_n) := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Here is an outline of the argument with some details being left to the reader.

1. If $f : X \to [0, \infty]$ is $\mathcal{M}$ -measurable, then
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ -measurable. Thus by the induction hypothesis, the right side of Eq. (9.17) is well defined.
2. If $\sigma \in S_n$ (the permutations of $\{1, 2, \ldots, n\}$) we may define a measure π on $(X, \mathcal{M})$ by;
$$\pi(A) := \int_{X_{\sigma 1}} d\mu_{\sigma 1}(x_{\sigma 1}) \ldots \int_{X_{\sigma n}} d\mu_{\sigma n}(x_{\sigma n})\, 1_A(x_1, \ldots, x_n). \tag{9.19}$$
It is easy to check that π is a measure which satisfies Eq. (9.16). Using the σ – finiteness assumptions and the fact that
$$\mathcal{P} := \{A_1 \times \cdots \times A_n : A_i \in \mathcal{M}_i \text{ for } 1 \leq i \leq n\}$$
is a π – system such that $\sigma(\mathcal{P}) = \mathcal{M}$, it follows from Exercise 5.1 that there is only one such measure satisfying Eq. (9.16). Thus the formula for π in Eq. (9.19) is independent of $\sigma \in S_n$.
3. From Eq. (9.19) and the usual simple function approximation arguments we may conclude that Eq. (9.17) is valid.
Now suppose that $f \in L^1(X, \mathcal{M}, \pi)$.
4. Using step 1 it is easy to check that
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \bar{\int}_{X_i} f(x_1, \ldots, x_i, \ldots, x_n)\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable. Indeed,
$$(x_1, \ldots, \hat{x}_i, \ldots, x_n) \to \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i)$$
is $\mathcal{M}^i$ – measurable and therefore
$$E := \left\{(x_1, \ldots, \hat{x}_i, \ldots, x_n) : \int_{X_i} |f(x_1, \ldots, x_i, \ldots, x_n)|\, d\mu_i(x_i) < \infty\right\} \in \mathcal{M}^i.$$
Now let $u := \operatorname{Re} f$ and $v := \operatorname{Im} f$ and $u_\pm$ and $v_\pm$ are the positive and negative parts of u and v respectively, then
$$\begin{aligned}\bar{\int}_{X_i} f(x)\, d\mu_i(x_i) &= \int_{X_i} 1_E(x^i)\, f(x)\, d\mu_i(x_i) \\ &= \int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) + i \int_{X_i} 1_E(x^i)\, v(x)\, d\mu_i(x_i).\end{aligned}$$
Both of these later terms are $\mathcal{M}^i$ – measurable since, for example,
$$\int_{X_i} 1_E(x^i)\, u(x)\, d\mu_i(x_i) = \int_{X_i} 1_E(x^i)\, u_+(x)\, d\mu_i(x_i) - \int_{X_i} 1_E(x^i)\, u_-(x)\, d\mu_i(x_i)$$
which is $\mathcal{M}^i$ – measurable by step 1.

5. It now follows by induction that the right side of Eq. (9.18) is well defined.
6. Let $i := \sigma n$ and $T : X \to X_i \times X^i$ be the obvious identification;
$$T\left(x_i, (x_1, \dots, \hat{x}_i, \dots, x_n)\right) = (x_1, \dots, x_n).$$
One easily verifies T is $\mathcal{M}/\mathcal{M}_i \otimes \mathcal{M}^i$ – measurable (use Corollary 6.19 repeatedly) and that $\pi \circ T^{-1} = \mu_i \otimes \mu^i$ (see Exercise 5.1).
7. Let $f \in L^1(\pi)$. Combining step 6. with the abstract change of variables Theorem (Exercise 7.11) implies
$$\int_X f d\pi = \int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right). \tag{9.20}$$
By Theorem 9.9, we also have
$$\begin{aligned}\int_{X_i \times X^i} (f \circ T)\, d\left(\mu_i \otimes \mu^i\right) &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f \circ T(x_i, x^i) \\ &= \int_{X^i} d\mu^i\left(x^i\right) \bar{\int}_{X_i} d\mu_i(x_i)\ f(x_1, \dots, x_n).\end{aligned} \tag{9.21}$$
Then by the induction hypothesis,
$$\int_{X^i} d\mu^i(x_i) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) = \prod_{j \neq i} \bar{\int}_{X_j} d\mu_j(x_j) \bar{\int}_{X_i} d\mu_i(x_i)\, f(x_1, \dots, x_n) \tag{9.22}$$
where the ordering the integrals in the last product are inconsequential. Combining Eqs. (9.20) – (9.22) completes the proof.

■

Convention: We are now going to drop the bar above the integral sign with the understanding that $\int_X f d\mu = 0$ whenever $f : X \to \mathbb{C}$ is a measurable function such that $\int_X |f|\, d\mu = \infty$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) non-integrable function we will interpret $\int_X f d\mu$ to be infinite.

Example 9.11. In this example we will show
$$\lim_{M \to \infty} \int_0^M \frac{\sin x}{x} dx = \pi/2. \tag{9.23}$$
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 9.12. We have
$$\int_0^\infty \frac{\sin x}{x} e^{-\Lambda x} dx = \frac{1}{2}\pi - \arctan \Lambda \text{ for all } \Lambda > 0 \tag{9.24}$$
and for$\Lambda, M \in [0, \infty)$,
$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C \frac{e^{-M\Lambda}}{M} \tag{9.25}$$
where $C = \max_{x \ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (9.23) 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 $x \neq 0$. Making use of the identity
$$\int_0^\infty e^{-tx} dt = 1/x$$
and Fubini's theorem,

$$\begin{aligned}\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x \, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\ &= \int_0^\infty dt \int_0^M dx \sin x \, e^{-(\Lambda+t)x} \\ &= \int_0^\infty \frac{1-(\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2+1} dt \\ &= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\ &= \frac{1}{2}\pi - \arctan \Lambda - \varepsilon(M,\Lambda) \qquad (9.26)\end{aligned}$$

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. (9.26) proves Eq. (9.25) from which Eq. (9.23) follows by taking $\Lambda \to \infty$ and Eq. (9.24) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

Lemma 9.13. *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 $x \in \mathbb{R}$,

$$\varphi(x) = \varphi(-M) + \int_{-M}^x \varphi'(y)\, dy. \qquad (9.27)$$

Under the stated assumptions on φ, we may use either the monotone or the dominated convergence theorem to let $M \to \infty$ in Eq. (9.27) 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 9.14. Here are a couple of examples involving Lemma 9.13.

1. Suppose X is a random variable, then
$$\mathbb{E}\left[e^X\right] = \int_{-\infty}^\infty P(X > y)\, e^y dy = \int_0^\infty P(X > \ln u)\, du, \qquad (9.28)$$
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
$$\mathbb{E}X^p = p \int_0^\infty y^{p-1} P(X > y)\, dy. \qquad (9.29)$$

9.4 Fubini's Theorem and Completions*

Notation 9.15 *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 9.16. *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,*

$$\left(x \to \int_Y f_x d\nu\right) \in L^1(\mu) \;\; and \;\; \left(y \to \int_X f^y d\mu\right) \in L^1(\nu)$$

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

$$0 = (\mu\otimes\nu)(E) = \int\limits_X \nu({}_xE)d\mu(x) = \int\limits_X \mu(E_y)d\nu(y).$$

This shows that

$$\mu(\{x : \nu({}_xE) \neq 0\}) = 0 \text{ and } \nu(\{y : \mu(E_y) \neq 0\}) = 0,$$

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 $\{(x,y) : h(x,y) \neq 0\} \subset E$ and $(\mu\otimes\nu)(E) = 0$. Therefore $|h(x,y)| \le 1_E(x,y)$ and $(\mu\otimes\nu)(E) = 0$. Since

$$\{h_x \neq 0\} = \{y \in Y : h(x,y) \neq 0\} \subset {}_xE \text{ and}$$
$$\{h_y \neq 0\} = \{x \in X : h(x,y) \neq 0\} \subset E_y$$

we learn that for μ a.e. x and ν a.e. y that $\{h_x \neq 0\} \in \mathcal{M}$, $\{h_y \neq 0\} \in \mathcal{N}$, $\nu(\{h_x \neq 0\}) = 0$ and a.e. and $\mu(\{h_y \neq 0\}) = 0$. 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

$$0 = \int_{X\times Y} h d\lambda = \int_Y \left(\int_X h d\mu\right) d\nu = \int_X \left(\int_Y h d\nu\right) d\mu.$$

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 9.17 *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 9.18. *A function $f : \mathbb{R}^d \to \mathbb{R}$ is* ***Lebesgue measurable*** *if $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{L}_d$.*

Notation 9.19 *I will often be sloppy in the sequel and write m for m^d and dx for $dm(x) = dm^d(x)$, i.e.*

$$\int_{\mathbb{R}^d} f(x)\, dx = \int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d} f dm^d.$$

Hopefully the reader will understand the meaning from the context.

Theorem 9.20. *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 9.13. ■

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 9.21 (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, $f : T(\Omega) \to [0, \infty]$,*

$$\int_\Omega f(T(x)) \,|\det T'(x)|\, dx = \int_{T(\Omega)} f(y)\, dy, \tag{9.30}$$

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

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \ldots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \ldots & \partial_d T_d(x) \end{bmatrix}, \tag{9.31}$$

i.e. the i - j – matrix entry of $T'(x)$ is given by $T'(x)_{ij} = \partial_i T_j(x)$ where $T(x) = (T_1(x), \ldots, T_d(x))^{\mathrm{tr}}$ and $\partial_i = \partial/\partial x_i$.

Remark 9.22. Theorem 9.21 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 $T(\Omega)$.

[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.

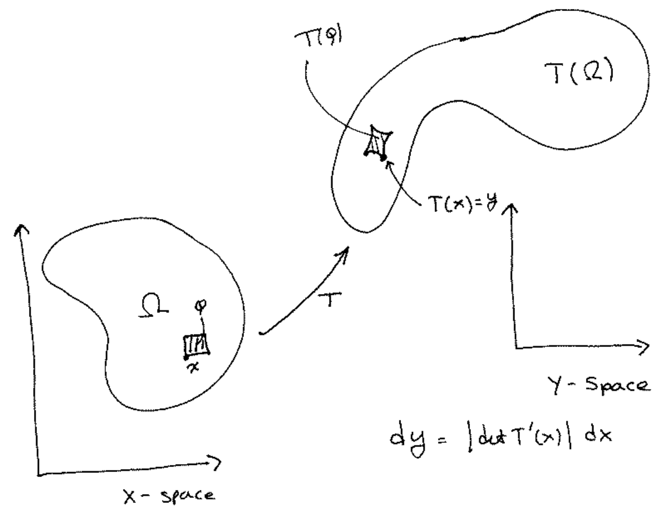


Fig. 9.1. The geometric setup of Theorem 9.21.

Note: you may skip the rest of this section!

Proof. The proof will be by induction on d. The case $d = 1$ was essentially done in Exercise 7.12. 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

$$\int_{[a,b]} f(T(x))\,|T'(x)|\,dx = \int_{T([a,b])} f(y)\,dy \tag{9.32}$$

whenever f is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 8.16 then implies that Eq. (9.32) 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 $n = 1, 2, \dots < N$ 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. (9.30) 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. (9.31).

Case 1. Suppose $T(x)$ has the form

$$T(x) = (x_i, T_2(x), \dots, T_d(x)) \tag{9.33}$$

or

$$T(x) = (T_1(x), \dots, T_{d-1}(x), x_i) \tag{9.34}$$

for some $i \in \{1, \dots, d\}$. For definiteness we will assume T is as in Eq. (9.33), the case of T in Eq. (9.34) 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

$$i_t(w) := w_t := (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}),$$

Ω_t be the (possibly empty) open subset of $\mathbb{R}^{d-1}$ defined by

$$\Omega_t := \left\{w \in \mathbb{R}^{d-1} : (w_1, \dots, w_{i-1}, t, w_{i+1}, \dots, w_{d-1}) \in \Omega\right\}$$

and $T_t : \Omega_t \to \mathbb{R}^{d-1}$ be defined by

$$T_t(w) = (T_2(w_t), \dots, T_d(w_t)),$$

see Figure 9.2. Expanding $\det T'(w_t)$ along the first row of the matrix $T'(w_t)$

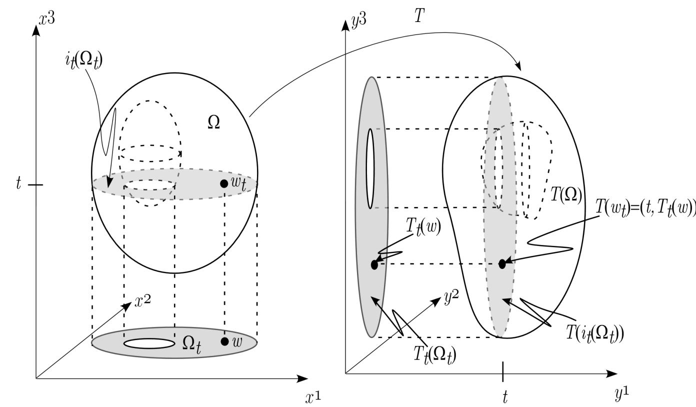


Fig. 9.2. In this picture $d = i = 3$ and Ω is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map T and slicing the set Ω along planes where $x_3 = t$.

shows

$$|\det T'(w_t)| = |\det T_t'(w)|.$$

Now by the Fubini-Tonelli Theorem and the induction hypothesis,

$$\begin{aligned}
\int_{\Omega} f\circ T|\det T'|dm &= \int_{\mathbb{R}^d} 1_\Omega \cdot f\circ T|\det T'|dm \\
&= \int_{\mathbb{R}^d} 1_\Omega\left(w_t\right)\left(f\circ T\right)\left(w_t\right)|\det T'\left(w_t\right)|dwdt \\
&= \int_{\mathbb{R}}\left[\int_{\Omega_t}\left(f\circ T\right)\left(w_t\right)|\det T'\left(w_t\right)|dw\right]dt \\
&= \int_{\mathbb{R}}\left[\int_{\Omega_t} f\left(t,T_t\left(w\right)\right)|\det T_t'\left(w\right)|dw\right]dt \\
&= \int_{\mathbb{R}}\left[\int_{T_t(\Omega_t)} f\left(t,z\right)dz\right]dt = \int_{\mathbb{R}}\left[\int_{\mathbb{R}^{d-1}} 1_{T(\Omega)}\left(t,z\right)f\left(t,z\right)dz\right]dt \\
&= \int_{T(\Omega)} f\left(y\right)dy
\end{aligned}$$

wherein the last two equalities we have used Fubini-Tonelli along with the identity;

$$T\left(\Omega\right) = \sum_{t\in\mathbb{R}} T\left(i_t\left(\Omega\right)\right) = \sum_{t\in\mathbb{R}}\left\{\left(t,z\right) : z\in T_t\left(\Omega_t\right)\right\}.$$

Case 2. (Eq. (9.30) 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

$$\int_W f\circ T|\det T'|dm = \int_{T(W)} fdm$$

holds for all Borel measurable function, $f:T(W)\to[0,\infty]$. Let M_i be the 1-i minor of $T'\left(x_0\right),$ i.e. the determinant of $T'\left(x_0\right)$ with the first row and i^{th} – column removed. Since

$$0\neq\det T'\left(x_0\right) = \sum_{i=1}^d\left(-1\right)^{i+1}\partial_i T_j\left(x_0\right)\cdot M_i,$$

there must be some i such that $M_i\neq 0$. Fix an i such that $M_i\neq 0$ and let,

$$S\left(x\right) := \left(x_i, T_2\left(x\right),\dots,T_d\left(x\right)\right). \tag{9.35}$$

Observe that $|\det S'\left(x_0\right)| = |M_i|\neq 0$. Hence by the inverse function Theorem, there exist an open neighborhood W of x_0 such that $W\subset_o\Omega$ and $S\left(W\right)\subset_o\mathbb{R}^d$ and $S:W\to S\left(W\right)$ is a C^1 – diffeomorphism. Let $R:S\left(W\right)\to T\left(W\right)\subset_o\mathbb{R}^d$ to be the C^1 – diffeomorphism defined by

$$R\left(z\right) := T\circ S^{-1}\left(z\right) \text{ for all } z\in S\left(W\right).$$

Because

$$\left(T_1\left(x\right),\dots,T_d\left(x\right)\right) = T\left(x\right) = R\left(S\left(x\right)\right) = R\left(\left(x_i,T_2\left(x\right),\dots,T_d\left(x\right)\right)\right)$$

for all $x\in W$, if

$$\left(z_1,z_2,\dots,z_d\right) = S\left(x\right) = \left(x_i,T_2\left(x\right),\dots,T_d\left(x\right)\right)$$

then

$$R\left(z\right) = \left(T_1\left(S^{-1}\left(z\right)\right),z_2,\dots,z_d\right). \tag{9.36}$$

Observe that S is a map of the form in Eq. (9.33), R is a map of the form in Eq. (9.34), $T'\left(x\right) = R'\left(S\left(x\right)\right)S'\left(x\right)$ (by the chain rule) and (by the multiplicative property of the determinant)

$$|\det T'\left(x\right)| = |\det R'\left(S\left(x\right)\right)||\det S'\left(x\right)| \;\forall\, x\in W.$$

So if $f:T(W)\to[0,\infty]$ is a Borel measurable function, two applications of the results in Case 1. shows,

$$\begin{aligned}
\int_W f\circ T\cdot|\det T'|dm &= \int_W\left(f\circ R\cdot|\det R'|\right)\circ S\cdot|\det S'|\; dm \\
&= \int_{S(W)} f\circ R\cdot|\det R'|dm = \int_{R(S(W))} fdm \\
&= \int_{T(W)} fdm
\end{aligned}$$

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

$$K_n := \left\{x\in\Omega : \operatorname{dist}(x,\Omega^c)\geq 1/n \text{ and } |x|\leq n\right\}.$$

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. (9.30) 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 $\tilde{W}_i := W_i\setminus\left(W_1\cup\dots\cup W_{i-1}\right)$ 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 9.23. When $d=1$, one often learns the change of variables formula as

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_{T(a)}^{T(b)} f(y)\,dy \tag{9.37}$$

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. (9.37) is implies Eq. (9.30) with $\Omega=(a,b)$. On the other hand if $T'<0$ on (a,b) then $T((a,b)) = (T(b),T(a))$ and Eq. (9.37) 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. (9.30). On the other hand Eq. (9.37) is more general than Eq. (9.30) since it does not require T to be injective. The standard proof of Eq. (9.37) 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,

$$\begin{aligned}\int_a^b f(T(x))\,T'(x)\,dx &= \int_a^b F'(T(x))\,T'(x)\,dx = \int_a^b \frac{d}{dx}[F(T(x))]\,dx \\ &= F(T(x))|_a^b = \int_{T(a)}^{T(b)} f(y)\,dy.\end{aligned}$$

An application of Dynkin's multiplicative systems theorem now shows that Eq. (9.37) 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).

Exercise 9.2. Continuing the setup in Theorem 9.21, 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. (9.30) holds.

Example 9.24. Continuing the setup in Theorem 9.21, if $A\in\mathcal{B}_\Omega$, then

$$\begin{aligned}m(T(A)) &= \int_{\mathbb{R}^d} 1_{T(A)}(y)\,dy = \int_{\mathbb{R}^d} 1_{T(A)}(Tx)\,|\det T'(x)|\,dx \\ &= \int_{\mathbb{R}^d} 1_A(x)\,|\det T'(x)|\,dx\end{aligned}$$

wherein the second equality we have made the change of variables, $y=T(x)$. Hence we have shown

$$d(m\circ T) = |\det T'(\cdot)|\,dm.$$

Taking $T\in GL(d,\mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d\times d$ invertible matrices in the previous example implies $m\circ T = |\det T|\,m$, i.e.

$$m(T(A)) = |\det T|\,m(A) \text{ for all } A\in\mathcal{B}_{\mathbb{R}^d}. \tag{9.38}$$

This equation also shows that $m\circ T$ and m have the same null sets and hence the equality in Eq. (9.38) is valid for any $A\in\mathcal{L}_d$. In particular we may conclude that m is invariant under those $T\in GL(d,\mathbb{R})$ with $|\det(T)|=1$. For example if T is a rotation (i.e. $T^{\mathrm{tr}}T=I$), then $\det T=\pm1$ and hence m is invariant under all rotations. This is not obvious from the definition of m^d as a product measure!

Example 9.25. Suppose that $T(x) = x+b$ for some $b\in\mathbb{R}^d$. In this case $T'(x) = I$ and therefore it follows that

$$\int_{\mathbb{R}^d} f(x+b)\,dx = \int_{\mathbb{R}^d} f(y)\,dy$$

for all measurable $f:\mathbb{R}^d\to[0,\infty]$ or for any $f\in L^1(m)$. In particular Lebesgue measure is invariant under translations.

Example 9.26 (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

$$dx = |\det T'(r,\theta)|\, dr d\theta = r dr d\theta.$$

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 9.21 that

$$\int_{\mathbb{R}^2} f(x)dx = \int_0^{2\pi} d\theta \int_0^\infty dr\ r\cdot f(r(\cos\theta,\sin\theta)) \tag{9.39}$$

for any Borel measurable function $f : \mathbb{R}^2 \to [0,\infty]$.

Example 9.27 (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

$$U_xV_y - U_yV_x = U_x^2 + V_x^2 = |f'|^2.$$

Therefore

$$dudv = |f'(x+iy)|^2\, dxdy.$$

Example 9.28. In this example we will evaluate the integral

$$I := \iint_\Omega \left(x^4 - y^4\right) dxdy$$

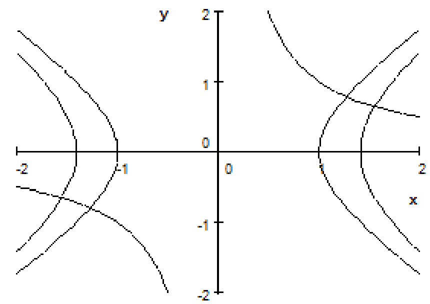


Fig. 9.3. The region Ω consists of the two curved rectangular regions shown.

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,

$$(u,v) := T(x,y) = \left(x^2 - y^2, xy\right),$$

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

$$I_\pm := \iint_{\pm D} \left(x^4 - y^4\right) dxdy = \frac{1}{2}\iint_{(1,2)\times(0,1)} ududv = \frac{1}{2}\frac{u^2}{2}\Big|_1^2 \cdot 1 = \frac{3}{4}$$

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

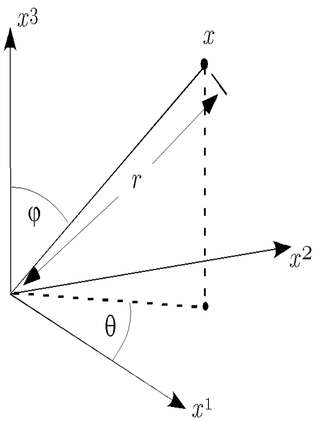


Fig. 9.4. The relation of x to (r, ϕ, θ) in spherical coordinates.

$$\begin{aligned} T(r, \varphi, \theta) &= (r \sin\varphi \cos\theta, r \sin\varphi \sin\theta, r\cos\varphi) \\ &= r(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi), \end{aligned}$$

see Figure 9.4. By making the change of variables $x = T(r, \varphi, \theta)$, show

$$\int_{\mathbb{R}^3} f(x)dx = \int_0^\pi d\varphi \int_0^{2\pi} d\theta \int_0^\infty dr\ r^2 \sin\varphi \cdot f(T(r, \varphi, \theta))$$

for any Borel measurable function, $f : \mathbb{R}^3 \to [0, \infty]$.

Lemma 9.29. *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,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \end{aligned} \tag{9.40}$$

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. (9.39), we find,

$$\begin{aligned} I_2(a) &= \int_0^\infty dr\ r \int_0^{2\pi} d\theta\ e^{-ar^2} = 2\pi \int_0^\infty r e^{-ar^2} dr \\ &= 2\pi \lim_{M\to\infty} \int_0^M r e^{-ar^2} dr = 2\pi \lim_{M\to\infty} \frac{e^{-ar^2}}{-2a} \int_0^M = \frac{2\pi}{2a} = \pi/a. \end{aligned}$$

This shows that $I_2(a) = \pi/a$ and the result now follows from Eq. (9.40). ■

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 9.30. *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

$$m((1, 1+\varepsilon]E) = m(E_{1+\varepsilon} \setminus E_1) = \left\{(1+\varepsilon)^d - 1\right\} m(E_1).$$

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 9.26 and Exercise 9.3.

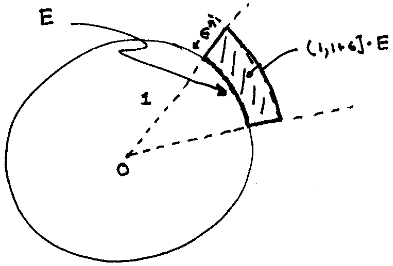


Fig. 9.5. Motivating the definition of surface measure for a sphere.

Theorem 9.31 (Polar Coordinates). *If $f : \mathbb{R}^d \to [0,\infty]$ is a $(\mathcal{B}_{R^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x)dm(x) = \int_{(0,\infty)\times S^{d-1}} f(r\omega)r^{d-1}\, drd\sigma(\omega). \tag{9.41}$$

In particular if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable then

$$\int_{\mathbb{R}^d} f(|x|)dx = \int_0^\infty f(r)dV(r) \tag{9.42}$$

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 7.11,

$$\int_{\mathbb{R}^d} fdm = \int_{\mathbb{R}^d\setminus\{0\}} \left(f\circ\Phi^{-1}\right)\circ\Phi\; dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right)\; d\left(\Phi_* m\right) \tag{9.43}$$

and therefore to prove Eq. (9.41) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right)\ \forall\ A\in\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}. \tag{9.44}$$

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,

$$\begin{aligned}(\Phi_* m)\left((a,b]\times E\right) &= m\left(bE_1\setminus aE_1\right) = m(bE_1) - m(aE_1)\\ &= b^d m(E_1) - a^d m(E_1) = d\cdot m(E_1)\int_a^b r^{d-1}dr.\end{aligned} \tag{9.45}$$

Letting $d\rho(r) = r^{d-1}dr$, i.e.

$$\rho(J) = \int_J r^{d-1}dr\ \forall\ J\in\mathcal{B}_{(0,\infty)}, \tag{9.46}$$

Eq. (9.45) may be written as

$$(\Phi_* m)\left((a,b]\times E\right) = \rho((a,b])\cdot\sigma(E) = (\rho\otimes\sigma)\left((a,b]\times E\right). \tag{9.47}$$

Since

$$\mathcal{E} = \{(a,b]\times E : 0<a<b \text{ and } E\in\mathcal{B}_{S^{d-1}}\},$$

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. (9.47) that $\Phi_* m = \rho\otimes\sigma$. Using this result in Eq. (9.43) gives

$$\int_{\mathbb{R}^d} fdm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right)\; d\left(\rho\otimes\sigma\right)$$

which combined with Tonelli's Theorem 9.6 proves Eq. (9.43). ■

Corollary 9.32. *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1}\subset\mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{9.48}$$

where Γ is the gamma function is as in Example 7.47 and 7.50.

Proof. Using Theorem 9.31 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

$$\begin{aligned}\int_0^\infty r^{d-1}e^{-r^2}dr &= \int_0^\infty u^{\frac{d-1}{2}}e^{-u}\frac{1}{2}u^{-1/2}du\\ &= \frac{1}{2}\int_0^\infty u^{\frac{d}{2}-1}e^{-u}du = \frac{1}{2}\Gamma(d/2).\end{aligned} \tag{9.49}$$

Combing the the last two equations with Lemma 9.29 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. (9.48). ■

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

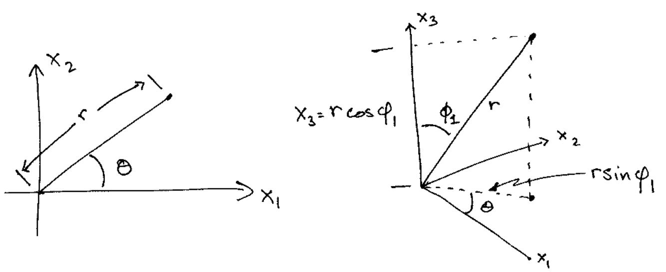


Fig. 9.6. Setting up polar coordinates in two and three dimensions.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} T_2(\theta, r\sin\varphi_1) \\ r\cos\varphi_1 \end{pmatrix} = \begin{pmatrix} r\sin\varphi_1\cos\theta \\ r\sin\varphi_1\sin\theta \\ r\cos\varphi_1 \end{pmatrix} =: T_3(\theta, \varphi_1, r,).$$

We continue to work inductively this way to define

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1},) \\ r\cos\varphi_{n-1} \end{pmatrix} = T_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r).$$

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,

$$\begin{aligned} x_1 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\cos\varphi_1 \\ &\vdots \\ x_{n-2} &= r\sin\varphi_{n-2}\sin\varphi_{n-3}\cos\varphi_{n-4} \\ x_{n-1} &= r\sin\varphi_{n-2}\cos\varphi_{n-3} \\ x_n &= r\cos\varphi_{n-2}. \end{aligned} \tag{9.50}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x)dm(x) \\ &= \int_0^\infty dr \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} d\varphi_1\dots d\varphi_{n-2}d\theta \begin{bmatrix} \Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) \\ \times f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)) \end{bmatrix} \end{aligned} \tag{9.51}$$

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 9.33. *The Jacobian, Δ_n is given by*

$$\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r) = r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1. \tag{9.52}$$

If f is a function on rS^{n-1} – the sphere of radius r centered at 0 inside of $\mathbb{R}^n$, then

$$\begin{aligned} &\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1}\int_{S^{n-1}} f(r\omega)d\sigma(\omega) \\ &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} f(T_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r))\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r)d\varphi_1\dots d\varphi_{n-2}d\theta \end{aligned} \tag{9.53}$$

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

$$\begin{aligned} &\Delta_{n+1}(\theta, \varphi_1, \dots, \varphi_{n-2}, \varphi_{n-1}, r) \\ &= \left|\begin{bmatrix} \frac{\partial T_n}{\partial\theta} & \frac{\partial T_n}{\partial\varphi_1} & \dots & \frac{\partial T_n}{\partial\varphi_{n-2}} & \frac{\partial T_n}{\partial\rho} r\cos\varphi_{n-1} & \frac{\partial T_n}{\partial\rho}\sin\varphi_{n-1} \\ 0 & 0 & \dots & 0 & -r\sin\varphi_{n-1} & \cos\varphi_{n-1} \end{bmatrix}\right| \\ &= r\left(\cos^2\varphi_{n-1} + \sin^2\varphi_{n-1}\right)\Delta_n(, \theta, \varphi_1, \dots, \varphi_{n-2}, \rho) \\ &= r\Delta_n(\theta, \varphi_1, \dots, \varphi_{n-2}, r\sin\varphi_{n-1}), \end{aligned}$$

i.e.

$$\Delta_{n+1}(\theta,\varphi_1,\dots,\varphi_{n-2},\varphi_{n-1},r)=r\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r\sin\varphi_{n-1}). \tag{9.54}$$

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 9.26, Eq. (9.54) 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. (9.52). Equation (9.53) now follows from Eqs. (9.41), (9.51) and (9.52). ■

As a simple application, Eq. (9.53) implies

$$\begin{aligned}\sigma(S^{n-1})&=\int_{0\le\varphi_i\le\pi,0\le\theta\le2\pi}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1d\varphi_1\dots d\varphi_{n-2}d\theta\\ &=2\pi\prod_{k=1}^{n-2}\gamma_k=\sigma(S^{n-2})\gamma_{n-2}\end{aligned} \tag{9.55}$$

where $\gamma_k:=\int_0^\pi\sin^k\varphi d\varphi$. If $k\ge1$, we have by integration by parts that,

$$\begin{aligned}\gamma_k&=\int_0^\pi\sin^k\varphi d\varphi=-\int_0^\pi\sin^{k-1}\varphi\ d\cos\varphi=2\delta_{k,1}+(k-1)\int_0^\pi\sin^{k-2}\varphi\ \cos^2\varphi d\varphi\\ &=2\delta_{k,1}+(k-1)\int_0^\pi\sin^{k-2}\varphi\ \left(1-\sin^2\varphi\right)d\varphi=2\delta_{k,1}+(k-1)\left[\gamma_{k-2}-\gamma_k\right]\end{aligned}$$

and hence γ_k satisfies $\gamma_0=\pi$, $\gamma_1=2$ and the recursion relation

$$\gamma_k=\frac{k-1}{k}\gamma_{k-2}\text{ for }k\ge2.$$

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. (9.55) may be written as

$$\sigma(S^n)=\sigma\left(S^{n-1}\right)\gamma_{n-1} \tag{9.56}$$

which combined with $\sigma\left(S^1\right)=2\pi$ implies

$$\begin{aligned}\sigma\left(S^1\right)&=2\pi,\\ \sigma(S^2)&=2\pi\cdot\gamma_1=2\pi\cdot2,\\ \sigma(S^3)&=2\pi\cdot2\cdot\gamma_2=2\pi\cdot2\cdot\frac{1}{2}\pi=\frac{2^2\pi^2}{2!!},\\ \sigma(S^4)&=\frac{2^2\pi^2}{2!!}\cdot\gamma_3=\frac{2^2\pi^2}{2!!}\cdot2\frac{2}{3}=\frac{2^3\pi^2}{3!!}\\ \sigma(S^5)&=2\pi\cdot2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi=\frac{2^3\pi^3}{4!!},\\ \sigma(S^6)&=2\pi\cdot2\cdot\frac{1}{2}\pi\cdot\frac{2}{3}2\cdot\frac{3}{4}\frac{1}{2}\pi\cdot\frac{4}{5}\frac{2}{3}2=\frac{2^4\pi^3}{5!!}\end{aligned}$$

and more generally that

$$\sigma(S^{2n})=\frac{2\left(2\pi\right)^n}{(2n-1)!!}\text{ and }\sigma(S^{2n+1})=\frac{(2\pi)^{n+1}}{(2n)!!} \tag{9.57}$$

which is verified inductively using Eq. (9.56). Indeed,

$$\sigma(S^{2n+1})=\sigma(S^{2n})\gamma_{2n}=\frac{2\left(2\pi\right)^n}{(2n-1)!!}\pi\frac{(2n-1)!!}{(2n)!!}=\frac{(2\pi)^{n+1}}{(2n)!!}$$

and

$$\sigma(S^{(n+1)})=\sigma(S^{2n+2})=\sigma(S^{2n+1})\gamma_{2n+1}=\frac{(2\pi)^{n+1}}{(2n)!!}2\frac{(2n)!!}{(2n+1)!!}=\frac{2\left(2\pi\right)^{n+1}}{(2n+1)!!}.$$

Using

$$(2n)!!=2n\left(2(n-1)\right)\dots(2\cdot1)=2^nn!$$

we may write $\sigma(S^{2n+1})=\frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (9.41) and (9.57 are in agreement. We may also write the formula in Eq. (9.57) 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 Gaussian Random Vectors

Definition 9.34 (Gaussian Random Vectors). *Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X : \Omega \to \mathbb{R}^d$ be a random vector. We say that X is Gaussian if there exists an $d \times d$ – symmetric matrix Q and a vector $\mu \in \mathbb{R}^d$ such that*

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}Q\lambda \cdot \lambda + i\mu \cdot \lambda\right) \text{ for all } \lambda \in \mathbb{R}^d. \tag{9.58}$$

We will write $X \stackrel{d}{=} N(Q, \mu)$ to denote a Gaussian random vector such that Eq. (9.58) holds.

Notice that if there exists a random variable satisfying Eq. (9.58) then its law is uniquely determined by Q and μ because of Corollary 8.11. In the exercises below your will develop some basic properties of Gaussian random vectors – see Theorem 9.38 for a summary of what you will prove.

Exercise 9.4. Show that Q must be non-negative in Eq. (9.58).

Definition 9.35. *Given a Gaussian random vector, X, we call the pair, (Q, μ) appearing in Eq. (9.58) the **characteristics** of X. We will also abbreviate the statement that X is a Gaussian random vector with characteristics (Q, μ) by writing $X \stackrel{d}{=} N(Q, \mu)$.*

Lemma 9.36. *Suppose that $X \stackrel{d}{=} N(Q, \mu)$ and $A : \mathbb{R}^d \to \mathbb{R}^m$ is a $m \times d$ – real matrix and $\alpha \in \mathbb{R}^m$, then $AX + \alpha \stackrel{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. In short we might abbreviate this by saying, $AN(Q, \mu) + \alpha \stackrel{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$.*

Proof. Let $\xi \in \mathbb{R}^m$, then

$$\begin{aligned}
\mathbb{E}\left[e^{i\xi \cdot (AX+\alpha)}\right] &= e^{i\xi \cdot \alpha}\mathbb{E}\left[e^{iA^{\mathrm{tr}}\xi \cdot X}\right] = e^{i\xi \cdot \alpha}\exp\left(-\frac{1}{2}QA^{\mathrm{tr}}\xi \cdot A^{\mathrm{tr}}\xi + i\mu \cdot A^{\mathrm{tr}}\xi\right) \\
&= e^{i\xi \cdot \alpha}\exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi \cdot \xi + iA\mu \cdot \xi\right) \\
&= \exp\left(-\frac{1}{2}AQA^{\mathrm{tr}}\xi \cdot \xi + i(A\mu + \alpha) \cdot \xi\right)
\end{aligned}$$

from which it follows that $AX + \alpha \stackrel{d}{=} N(AQA^{\mathrm{tr}}, A\mu + \alpha)$. ∎

Exercise 9.5. Let P be the probability measure on $\Omega := \mathbb{R}^d$ defined by

$$dP(x) := \left(\frac{1}{2\pi}\right)^{d/2} e^{-\frac{1}{2}x \cdot x}dx = \prod_{i=1}^{d}\left(\frac{1}{\sqrt{2\pi}}e^{-x_i^2/2}dx_i\right).$$

Show that $N : \Omega \to \mathbb{R}^d$ defined by $N(x) = x$ is Gaussian and satisfies Eq. (9.58) with $Q = I$ and $\mu = 0$. Also show

$$\mu_i = \mathbb{E}N_i \text{ and } \delta_{ij} = \mathrm{Cov}(N_i, N_j) \text{ for all } 1 \le i, j \le d. \tag{9.59}$$

Hint: use Exercise 7.15 and (of course) Fubini's theorem.

Exercise 9.6. Let A be any real $m \times d$ matrix and $\mu \in \mathbb{R}^m$ and set $X := AN + \mu$ where $\Omega = \mathbb{R}^d$, P, and N are as in Exercise 9.5. Show that X is Gaussian by showing Eq. (9.58) holds with $Q = AA^{\mathrm{tr}}$ (A^{tr} is the transpose of the matrix A) and $\mu = \mu$. Also show that

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \mathrm{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{9.60}$$

Remark 9.37 (Spectral Theorem). Recall that if Q is a real symmetric $d \times d$ matrix, then the spectral theorem asserts there exists an orthonormal basis, $\{u\}_{j=1}^d$, such that $Qu_j = \lambda_j u_j$ for some $\lambda_j \in \mathbb{R}$. Moreover, $\lambda_j \ge 0$ for all j is equivalent to Q being non-negative. When $Q \ge 0$ we may define $Q^{1/2}$ by

$$Q^{1/2}u_j := \sqrt{\lambda_j}u_j \text{ for } 1 \le j \le d.$$

Notice that $Q^{1/2} \ge 0$ and $Q = \left(Q^{1/2}\right)^2$ and $Q^{1/2}$ is still symmetric. If Q is positive definite, we may also define, $Q^{-1/2}$ by

$$Q^{-1/2}u_j := \frac{1}{\sqrt{\lambda_j}}u_j \text{ for } 1 \le j \le d$$

so that $Q^{-1/2} = \left[Q^{1/2}\right]^{-1}$.

Exercise 9.7. Suppose that Q is a positive definite (for simplicity) $d \times d$ real matrix and $\mu \in \mathbb{R}^d$ and let $\Omega = \mathbb{R}^d$, P, and N be as in Exercise 9.5. By Exercise 9.6 we know that $X = Q^{1/2}N + \mu$ is a Gaussian random vector satisfying Eq. (9.58). Use the multi-dimensional change of variables formula to show

$$\mathrm{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}}\exp\left(-\frac{1}{2}Q^{-1}(y-\mu) \cdot (y-\mu)\right)dy.$$

Let us summarize some of what the preceding exercises have shown.

Theorem 9.38. *To each positive definite $d \times d$ real symmetric matrix Q and $\mu \in \mathbb{R}^d$ there exist Gaussian random vectors, X, satisfying Eq. (9.58). Moreover for such an X,*

$$\mathrm{Law}_P(X)(dy) = \frac{1}{\sqrt{\det(2\pi Q)}}\exp\left(-\frac{1}{2}Q^{-1}(y-\mu) \cdot (y-\mu)\right)dy$$

where Q and μ may be computed from X using,

$$\mu_i = \mathbb{E}X_i \text{ and } Q_{ij} = \mathrm{Cov}(X_i, X_j) \text{ for all } 1 \le i, j \le m. \tag{9.61}$$

When Q is degenerate, i.e. $\operatorname{Nul}(Q) \neq \{0\}$, then $X = Q^{1/2}N + \mu$ is still a Gaussian random vectors satisfying Eq. (9.58). However now the $\operatorname{Law}_P(X)$ is a measure on $\mathbb{R}^d$ which is concentrated on the non-trivial subspace, $\operatorname{Nul}(Q)^{\perp}$ – the details of this are left to the reader for now.

Exercise 9.8 (Gaussian random vectors are "highly" integrable.). Suppose that $X : \Omega \to \mathbb{R}^d$ is a Gaussian random vector, say $X \overset{d}{=} N(Q, \mu)$. Let $\|x\| := \sqrt{x \cdot x}$ and $m := \max\{Qx \cdot x : \|x\| = 1\}$ be the largest eigenvalue[3] of Q. Then $\mathbb{E}\left[e^{\varepsilon \|X\|^2}\right] < \infty$ for every $\varepsilon < \frac{1}{2m}$.

Because of Eq. (9.61), for all $\lambda \in \mathbb{R}^d$ we have

$$\mu \cdot \lambda = \sum_{i=1}^{d} \mathbb{E}X_i \cdot \lambda_i = \mathbb{E}(\lambda \cdot X)$$

and

$$\begin{aligned} Q\lambda \cdot \lambda &= \sum_{i,j} Q_{ij}\lambda_i\lambda_j = \sum_{i,j} \lambda_i\lambda_j \operatorname{Cov}(X_i, X_j) \\ &= \operatorname{Cov}\left(\sum_i \lambda_i X_i, \sum_j \lambda_j X_j\right) = \operatorname{Var}(\lambda \cdot X). \end{aligned}$$

Therefore we may reformulate the definition of a Gaussian random vector as follows.

Definition 9.39 (Gaussian Random Vectors). *Let $(\Omega, \mathcal{B}, P)$ be a probability space. A random vector, $X : \Omega \to \mathbb{R}^d$, is Gaussian iff for all $\lambda \in \mathbb{R}^d$,*

$$\mathbb{E}\left[e^{i\lambda \cdot X}\right] = \exp\left(-\frac{1}{2}\operatorname{Var}(\lambda \cdot X) + i\mathbb{E}(\lambda \cdot X)\right). \tag{9.62}$$

In short, X is a Gaussian random vector iff $\lambda \cdot X$ is a Gaussian random variable for all $\lambda \in \mathbb{R}^d$.

Remark 9.40. To conclude that a random vector, $X : \Omega \to \mathbb{R}^d$, is Gaussian it is **not** enough to check that each of its components, $\{X_i\}_{i=1}^d$, are Gaussian random variables. The following simple counter example was provided by Nate Eldredge. Let $(X, Y) : \Omega \to \mathbb{R}^2$ be a Random vector such that $(X, Y)_* P = \mu \otimes \nu$ where $d\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$ and $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$. Then $(X, YX) : \Omega \to \mathbb{R}^2$ is a random vector such that both components, X and YX, are Gaussian random variables but (X, YX) is **not** a Gaussian random vector.

Exercise 9.9. Prove the assertion made in Remark 9.40. **Hint:** explicitly compute $\mathbb{E}\left[e^{i(\lambda_1 X + \lambda_2 XY)}\right]$.

[3] For those who know about operator norms observe that $m = \|Q\|$ in this case.

9.9 Kolmogorov's Extension Theorems

In this section we will extend the results of Section 5.5 to spaces which are not simply products of discrete spaces. We begin with a couple of results involving the topology on $\mathbb{R}^N$.

9.9.1 Regularity and compactness results

Theorem 9.41 (Inner-Outer Regularity). *Suppose μ is a probability measure on $\left(\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N}\right)$, then for all $B \in \mathcal{B}_{\mathbb{R}^N}$ we have*

$$\mu(B) = \inf\{\mu(V) : B \subset V \text{ and } V \text{ is open}\} \tag{9.63}$$

and

$$\mu(B) = \sup\{\mu(K) : K \subset B \text{ with } K \text{ compact}\}. \tag{9.64}$$

Proof. In this proof, C, and C_i will always denote a closed subset of $\mathbb{R}^N$ and V, V_i will always be open subsets of $\mathbb{R}^N$. Let $\mathcal{F}$ be the collection of sets, $A \in \mathcal{B}$, such that for all $\varepsilon > 0$ there exists an open set V and a closed set, C, such that $C \subset A \subset V$ and $\mu(V \setminus C) < \varepsilon$. The key point of the proof is to show $\mathcal{F} = \mathcal{B}$ for this certainly implies Equation (9.63) and also that

$$\mu(B) = \sup\{\mu(C) : C \subset B \text{ with } C \text{ closed}\}. \tag{9.65}$$

Moreover, by MCT, we know that if C is closed and $K_n := C \cap \left\{x \in \mathbb{R}^N : |x| \le n\right\}$, then $\mu(K_n) \uparrow \mu(C)$. This observation along with Eq. (9.65) shows Eq. (9.64) is valid as well.

To prove $\mathcal{F} = \mathcal{B}$, it suffices to show $\mathcal{F}$ is a σ – algebra which contains all closed subsets of $\mathbb{R}^N$. To the prove the latter assertion, given a closed subset, $C \subset \mathbb{R}^N$, and $\varepsilon > 0$, let

$$C_\varepsilon := \cup_{x \in C} B(x, \varepsilon)$$

where $B(x, \varepsilon) := \left\{y \in \mathbb{R}^N : |y - x| < \varepsilon\right\}$. Then C_ε is an open set and $C_\varepsilon \downarrow C$ as $\varepsilon \downarrow 0$. (You prove.) Hence by the DCT, we know that $\mu(C_\varepsilon \setminus C) \downarrow 0$ form which it follows that $C \in \mathcal{F}$.

We will now show that $\mathcal{F}$ is an algebra. Clearly $\mathcal{F}$ contains the empty set and if $A \in \mathcal{F}$ with $C \subset A \subset V$ and $\mu(V \setminus C) < \varepsilon$, then $V^c \subset A^c \subset C^c$ with $\mu(C^c \setminus V^c) = \mu(V \setminus C) < \varepsilon$. This shows $A^c \in \mathcal{F}$. Similarly if $A_i \in \mathcal{F}$ for $i = 1, 2$ and $C_i \subset A_i \subset V_i$ with $\mu(V_i \setminus C_i) < \varepsilon$, then

$$C := C_1 \cup C_2 \subset A_1 \cup A_2 \subset V_1 \cup V_2 =: V$$

and

$$\mu(V \setminus C) \le \mu(V_1 \setminus C) + \mu(V_2 \setminus C) \\ \le \mu(V_1 \setminus C_1) + \mu(V_2 \setminus C_2) < 2\varepsilon.$$

This implies that $A_1 \cup A_2 \in \mathcal{F}$ and we have shown $\mathcal{F}$ is an algebra.

We now show that $\mathcal{F}$ is a σ – algebra. To do this it suffices to show $A := \sum_{n=1}^{\infty} A_n \in \mathcal{F}$ if $A_n \in \mathcal{F}$ with $A_n \cap A_m = \emptyset$ for $m \neq n$. Let $C_n \subset A_n \subset V_n$ with $\mu(V_n \setminus C_n) < \varepsilon 2^{-n}$ for all n and let $C^N := \cup_{n \le N} C_n$ and $V := \cup_{n=1}^{\infty} V_n$. Then $C^N \subset A \subset V$ and

$$\mu\left(V \setminus C^N\right) \le \sum_{n=0}^{\infty} \mu\left(V_n \setminus C^N\right) \le \sum_{n=0}^{N} \mu(V_n \setminus C_n) + \sum_{n=N+1}^{\infty} \mu(V_n) \\ \le \sum_{n=0}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \left[\mu(A_n) + \varepsilon 2^{-n}\right] \\ = \varepsilon + \sum_{n=N+1}^{\infty} \mu(A_n).$$

The last term is less that 2ε for N sufficiently large because $\sum_{n=1}^{\infty} \mu(A_n) = \mu(A) < \infty$. ■

Notation 9.42 *Let $I := [0,1]$, $Q = I^{\mathbb{N}}$, $\pi_j : Q \to I$ be the projection map, $\pi_j(x) = x_j$ (where $x = (x_1, x_2, \dots, x_j, \dots)$ for all $j \in \mathbb{N}$, and $\mathcal{B}_Q := \sigma(\pi_j : j \in \mathbb{N})$ be the product σ – algebra on Q. Let us further say that a sequence $\{x(m)\}_{m=1}^{\infty} \subset Q$, where $x(m) = (x_1(m), x_2(m), \dots)$, converges to $x \in Q$ iff $\lim_{m\to\infty} x_j(m) = x_j$ for all $j \in \mathbb{N}$. (This is just pointwise convergence.)*

Lemma 9.43 (Baby Tychonoff's Theorem). *The infinite dimensional cube, Q, is compact, i.e. every sequence $\{x(m)\}_{m=1}^{\infty} \subset Q$ has a convergent subsequence,$\{x(m_k)\}_{k=1}^{\infty}$.*

Proof. Since I is compact, it follows that for each $j \in \mathbb{N}$, $\{x_j(m)\}_{m=1}^{\infty}$ has a convergent subsequence. It now follow by Cantor's diagonalization method, that there is a subsequence, $\{m_k\}_{k=1}^{\infty}$, of $\mathbb{N}$ such that $\lim_{k\to\infty} x_j(m_k) \in I$ exists for all $j \in \mathbb{N}$. ■

Corollary 9.44 (Finite Intersection Property). *Suppose that $K_m \subset Q$ are sets which are, (i) closed under taking sequential limits[4], and (ii) have the finite intersection property, (i.e. $\cap_{m=1}^{n} K_m \neq \emptyset$ for all $m \in \mathbb{N}$), then $\cap_{m=1}^{\infty} K_m \neq \emptyset$.*

Proof. By assumption, for each $n \in \mathbb{N}$, there exists $x(n) \in \cap_{m=1}^{n} K_m$. Hence by Lemma 9.43 there exists a subsequence, $x(n_k)$, such that $x := \lim_{k\to\infty} x(n_k)$ exists in Q. Since $x(n_k) \in \cap_{m=1}^{n} K_m$ for all k large, and each K_m is closed under sequential limits, it follows that $x \in K_m$ for all m. Thus we have shown, $x \in \cap_{m=1}^{\infty} K_m$ and hence $\cap_{m=1}^{\infty} K_m \neq \emptyset$. ■

[4] For example, if $K_m = K_m' \times Q$ with K_m' being a closed subset of I^m, then K_m is closed under sequential limits.

9.9.2 Kolmogorov's Extension Theorem and Infinite Product Measures

Theorem 9.45 (Kolmogorov's Extension Theorem). *Let $I := [0,1]$. For each $n \in \mathbb{N}$, let μ_n be a probability measure on $(I^n, \mathcal{B}_{I^n})$ such that $\mu_{n+1}(A \times I) = \mu_n(A)$. Then there exists a unique measure, P on $(Q, \mathcal{B}_Q)$ such that*

$$P(A \times Q) = \mu_n(A) \tag{9.66}$$

for all $A \in \mathcal{B}_{I^n}$ and $n \in \mathbb{N}$.

Proof. Let $\mathcal{A} := \cup \mathcal{B}_n$ where $\mathcal{B}_n := \{A \times Q : A \in \mathcal{B}_{I^n}\} = \sigma(\pi_1, \dots, \pi_n)$, where $\pi_i(x) = x_i$ if $x = (x_1, x_2, \dots) \in Q$. Then define P on $\mathcal{A}$ by Eq. (9.66) which is easily seen (Exercise 9.10) to be a well defined finitely additive measure on $\mathcal{A}$. So to finish the proof it suffices to show if $B_n \in \mathcal{A}$ is a decreasing sequence such that

$$\inf_n P(B_n) = \lim_{n\to\infty} P(B_n) = \varepsilon > 0,$$

then $B := \cap B_n \neq \emptyset$.

To simplify notation, we may reduce to the case where $B_n \in \mathcal{B}_n$ for all n. To see this is permissible, Let us choose $1 \le n_1 < n_2 < n_3 < \dots$. such that $B_k \in \mathcal{B}_{n_k}$ for all k. (This is possible since $\mathcal{B}_n$ is increasing in n.) We now define a new decreasing sequence of sets, $\left\{\tilde{B}_k\right\}_{k=1}^{\infty}$ as follows,

$$\left(\tilde{B}_1, \tilde{B}_2, \dots\right) = \left(\overbrace{Q, \dots, Q}^{n_1 - 1 \text{ times}}, \overbrace{B_1, \dots, B_1}^{n_2 - n_1 \text{ times}}, \overbrace{B_2, \dots, B_2}^{n_3 - n_2 \text{ times}}, \overbrace{B_3, \dots, B_3}^{n_4 - n_3 \text{ times}}, \dots\right).$$

We then have $\tilde{B}_n \in \mathcal{B}_n$ for all n, $\lim_{n\to\infty} P\left(\tilde{B}_n\right) = \varepsilon > 0$, and $B = \cap_{n=1}^{\infty} \tilde{B}_n$. Hence we may replace B_n by $\tilde{B}_n$ if necessary so as to have $B_n \in \mathcal{B}_n$ for all n.

Since $B_n \in \mathcal{B}_n$, there exists $B_n' \in \mathcal{B}_{I^n}$ such that $B_n = B_n' \times Q$ for all n. Using the regularity Theorem 9.41, there are compact sets, $K_n' \subset B_n' \subset I^n$, such that $\mu_n(B_n' \setminus K_n') \le \varepsilon 2^{-n-1}$ for all $n \in \mathbb{N}$. Let $K_n := K_n' \times Q$, then $P(B_n \setminus K_n) \le \varepsilon 2^{-n-1}$ for all n. Moreover,

$$P(B_n \setminus [\cap_{m=1}^{n} K_m]) = P(\cup_{m=1}^{n} [B_n \setminus K_m]) \le \sum_{m=1}^{n} P(B_n \setminus K_m) \\ \le \sum_{m=1}^{n} P(B_m \setminus K_m) \le \sum_{m=1}^{n} \varepsilon 2^{-m-1} \le \varepsilon/2.$$

So, for all $n \in \mathbb{N}$,

$$P\left(\cap_{m=1}^{n} K_m\right) = P\left(B_n\right) - P\left(B_n \setminus \left[\cap_{m=1}^{n} K_m\right]\right) \geq \varepsilon - \varepsilon/2 = \varepsilon/2,$$

and in particular, $\cap_{m=1}^{n} K_m \neq \emptyset$. An application of Corollary 9.44 now implies, $\emptyset \neq \cap_n K_n \subset \cap_n B_n$. ■

Exercise 9.10. Show that Eq. (9.66) defines a well defined finitely additive measure on $\mathcal{A} := \cup \mathcal{B}_n$.

The next result is an easy corollary of Theorem 9.45.

Theorem 9.46. *Suppose $\{(X_n, \mathcal{M}_n)\}_{n\in\mathbb{N}}$ are standard Borel spaces (see Appendix 9.10 below), $X := \prod_{n\in\mathbb{N}} X_n$, $\pi_n : X \to X_n$ be the n^{th} – projection map, $\mathcal{B}_n := \sigma(\pi_k : k \leq n)$, $\mathcal{B} = \sigma(\pi_n : n \in \mathbb{N})$, and $T_n := X_{n+1} \times X_{n+2} \times \dots$. Further suppose that for each $n \in \mathbb{N}$ we are given a probability measure, μ_n on $\mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$ such that*

$$\mu_{n+1}\left(A \times X_{n+1}\right) = \mu_n\left(A\right) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n.$$

Then there exists a unique probability measure, P, on $(X, \mathcal{B})$ such that $P\left(A \times T_n\right) = \mu_n\left(A\right)$ for all $A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$.

Proof. Since each $(X_n, \mathcal{M}_n)$ is measure theoretic isomorphic to a Borel subset of I, we may assume that $X_n \in \mathcal{B}_I$ and $\mathcal{M}_n = (\mathcal{B}_I)_{X_n}$ for all n. Given $A \in \mathcal{B}_{I^n}$, let $\bar{\mu}_n\left(A\right) := \mu_n\left(A \cap \left[X_1 \times \dots \times X_n\right]\right)$ – a probability measure on $\mathcal{B}_{I^n}$. Furthermore,

$$\begin{aligned}\bar{\mu}_{n+1}\left(A \times I\right) &= \mu_{n+1}\left(\left[A \times I\right] \cap \left[X_1 \times \dots \times X_{n+1}\right]\right) \\ &= \mu_{n+1}\left(\left(A \cap \left[X_1 \times \dots \times X_n\right]\right) \times X_{n+1}\right) \\ &= \mu_n\left(\left(A \cap \left[X_1 \times \dots \times X_n\right]\right)\right) = \bar{\mu}_n\left(A\right).\end{aligned}$$

Hence by Theorem 9.45, there is a unique probability measure, $\bar{P}$, on $I^{\mathbb{N}}$ such that

$$\bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n\left(A\right) \text{ for all } n \in \mathbb{N} \text{ and } A \in \mathcal{B}_{I^n}.$$

We will now check that $P := \bar{P}|_{\otimes_{n=1}^{\infty}\mathcal{M}_n}$ is the desired measure. First off we have

$$\begin{aligned}\bar{P}\left(X\right) &= \lim_{n\to\infty} \bar{P}\left(X_1 \times \dots \times X_n \times I^{\mathbb{N}}\right) = \lim_{n\to\infty} \bar{\mu}_n\left(X_1 \times \dots \times X_n\right) \\ &= \lim_{n\to\infty} \mu_n\left(X_1 \times \dots \times X_n\right) = 1.\end{aligned}$$

Secondly, if $A \in \mathcal{M}_1 \otimes \dots \otimes \mathcal{M}_n$, we have

$$\begin{aligned}P\left(A \times T_n\right) &= \bar{P}\left(A \times T_n\right) = \bar{P}\left(\left(A \times I^{\mathbb{N}}\right) \cap X\right) \\ &= \bar{P}\left(A \times I^{\mathbb{N}}\right) = \bar{\mu}_n\left(A\right) = \mu_n\left(A\right).\end{aligned}$$

■

Here is an example of this theorem in action.

Theorem 9.47 (Infinite Product Measures). *Suppose that $\{\nu_n\}_{n=1}^{\infty}$ are a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\mathcal{B} := \otimes_{n\in\mathbb{N}}\mathcal{B}_{\mathbb{R}}$ is the product σ – algebra on $\mathbb{R}^{\mathbb{N}}$. Then there exists a unique probability measure, ν, on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$, such that*

$$\nu\left(A_1 \times A_2 \times \dots \times A_n \times \mathbb{R}^{\mathbb{N}}\right) = \nu_1\left(A_1\right) \dots \nu_n\left(A_n\right) \ \forall \ A_i \in \mathcal{B}_{\mathbb{R}} \ \& \ n \in \mathbb{N}. \tag{9.67}$$

Moreover, this measure satisfies,

$$\int_{\mathbb{R}^{\mathbb{N}}} f\left(x_1, \dots, x_n\right) d\nu\left(x\right) = \int_{\mathbb{R}^n} f\left(x_1, \dots, x_n\right) d\nu_1\left(x_1\right) \dots d\nu_n\left(x_n\right) \tag{9.68}$$

for all $n \in \mathbb{N}$ and $f : \mathbb{R}^n \to \mathbb{R}$ which are bounded and measurable or non-negative and measurable.

Proof. The measure ν is created by apply Theorem 9.46 with $\mu_n := \nu_1 \otimes \dots \otimes \nu_n$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n} = \otimes_{k=1}^{n}\mathcal{B}_{\mathbb{R}})$ for each $n \in \mathbb{N}$. Observe that

$$\mu_{n+1}\left(A \times \mathbb{R}\right) = \mu_n\left(A\right) \cdot \nu_{n+1}\left(\mathbb{R}\right) = \mu_n\left(A\right),$$

so that $\{\mu_n\}_{n=1}^{\infty}$ satisfies the needed consistency conditions. Thus there exists a unique measure ν on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$ such that

$$\nu\left(A \times \mathbb{R}^{\mathbb{N}}\right) = \mu_n\left(A\right) \text{ for all } A \in \mathcal{B}_{\mathbb{R}^n} \text{ and } n \in \mathbb{N}.$$

Taking $A = A_1 \times A_2 \times \dots \times A_n$ with $A_i \in \mathcal{B}_{\mathbb{R}}$ then gives Eq. (9.67). For this measure, it follows that Eq. (9.68) holds when $f = 1_{A_1 \times \dots \times A_n}$. Thus by an application of Theorem 8.2 with $\mathbb{M} = \{1_{A_1 \times \dots \times A_n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$ and $\mathbb{H}$ being the set of bounded measurable functions, $f : \mathbb{R}^n \to \mathbb{R}$, for which Eq. (9.68) shows that Eq. (9.68) holds for all bounded and measurable functions, $f : \mathbb{R}^n \to \mathbb{R}$. The statement involving non-negative functions follows by a simple limiting argument involving the MCT. ■

It turns out that the existence of infinite product measures require no topological restrictions on the measure spaces involved. See Theorem **??** below.

9.10 Appendix: Standard Borel Spaces*

For more information along the lines of this section, see Royden [12].

Definition 9.48. *Two measurable spaces, $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are said to be **isomorphic** if there exists a bijective map, $f : X \to Y$ such that $f(\mathcal{M}) = \mathcal{N}$ and $f^{-1}(\mathcal{N}) = \mathcal{M}$, i.e. both f and f^{-1} are measurable. In this case we say f is a measure theoretic isomorphism and we will write $X \cong Y$.*

Definition 9.49. *A measurable space, $(X, \mathcal{M})$ is said to be a **standard Borel space** if $(X, \mathcal{M}) \cong (B, \mathcal{B}_B)$ where B is a Borel subset of $\left((0,1), \mathcal{B}_{(0,1)}\right)$.*

Definition 9.50 (Polish spaces). *A **Polish space** is a separable topological space (X, τ) which admits a complete metric, ρ, such that $\tau = \tau_\rho$.*

The main goal of this chapter is to prove every Borel subset of a Polish space is a standard Borel space, see Corollary 9.60 below. Along the way we will show a number of spaces, including $[0,1]$, $(0,1]$, $[0,1]^d$, $\mathbb{R}^d$, and $\mathbb{R}^{\mathbb{N}}$, are all (measure theoretic) isomorphic to $(0,1)$. Moreover we also will see that the a countable product of standard Borel spaces is again a standard Borel space, see Corollary 9.57.

On first reading, you may wish to skip the rest of this section.

Lemma 9.51. *Suppose $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are measurable spaces such that $X = \sum_{n=1}^\infty X_n$, $Y = \sum_{n=1}^\infty Y_n$, with $X_n \in \mathcal{M}$ and $Y_n \in \mathcal{N}$. If $(X_n, \mathcal{M}_{X_n})$ is isomorphic to $(Y_n, \mathcal{N}_{Y_n})$ for all n then $X \cong Y$. Moreover, if $(X_n, \mathcal{M}_n)$ and $(Y_n, \mathcal{N}_n)$ are isomorphic measure spaces, then $(X := \prod_{n=1}^\infty X_n, \otimes_{n=1}^\infty \mathcal{M}_n)$ are $(Y := \prod_{n=1}^\infty Y_n, \otimes_{n=1}^\infty \mathcal{N}_n)$ are isomorphic.*

Proof. For each $n \in \mathbb{N}$, let $f_n : X_n \to Y_n$ be a measure theoretic isomorphism. Then define $f : X \to Y$ by $f = f_n$ on X_n. Clearly, $f : X \to Y$ is a bijection and if $B \in \mathcal{N}$, then

$$f^{-1}(B) = \cup_{n=1}^\infty f^{-1}(B \cap Y_n) = \cup_{n=1}^\infty f_n^{-1}(B \cap Y_n) \in \mathcal{M}.$$

This shows f is measurable and by similar considerations, f^{-1} is measurable as well. Therefore, $f : X \to Y$ is the desired measure theoretic isomorphism.

For the second assertion, let $f_n : X_n \to Y_n$ be a measure theoretic isomorphism of all $n \in \mathbb{N}$ and then define

$$f(x) = (f_1(x_1), f_2(x_2), \dots) \text{ with } x = (x_1, x_2, \dots) \in X.$$

Again it is clear that f is bijective and measurable, since

$$f^{-1}\left(\prod_{n=1}^\infty B_n\right) = \prod_{n=1}^\infty f_n^{-1}(B_n) \in \otimes_{n=1}^\infty \mathcal{N}_n$$

for all $B_n \in \mathcal{M}_n$ and $n \in \mathbb{N}$. Similar reasoning shows that f^{-1} is measurable as well. ■

Proposition 9.52. *Let $-\infty < a < b < \infty$. The following measurable spaces equipped with there Borel σ – algebras are all isomorphic; $(0,1)$, $[0,1]$, $(0,1]$, $[0,1)$, (a,b), $[a,b]$, $(a,b]$, $[a,b)$, $\mathbb{R}$, and $(0,1) \cup \Lambda$ where Λ is a finite or countable subset of $\mathbb{R} \setminus (0,1)$.*

Proof. It is easy to see by that any bounded open, closed, or half open interval is isomorphic to any other such interval using an affine transformation. Let us now show $(-1,1) \cong [-1,1]$. To prove this it suffices, by Lemma 9.51,to observe that

$$(-1,1) = \{0\} \cup \sum_{n=0}^\infty \left((-2^{-n}, -2^{-n}] \cup [2^{-n-1}, 2^{-n})\right)$$

and

$$[-1,1] = \{0\} \cup \sum_{n=0}^\infty \left([-2^{-n}, -2^{-n-1}) \cup (2^{-n-1}, 2^{-n}]\right).$$

Similarly $(0,1)$ is isomorphic to $(0,1]$ because

$$(0,1) = \sum_{n=0}^\infty [2^{-n-1}, 2^{-n}) \text{ and } (0,1] = \sum_{n=0}^\infty (2^{-n-1}, 2^{-n}].$$

The assertion involving $\mathbb{R}$ can be proved using the bijection, $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$.

If $\Lambda = \{1\}$, then by Lemma 9.51 and what we have already proved, $(0,1) \cup \{1\} = (0,1] \cong (0,1)$. Similarly if $N \in \mathbb{N}$ with $N \geq 2$ and $\Lambda = \{2, \dots, N+1\}$, then

$$(0,1) \cup \Lambda \cong (0,1] \cup \Lambda = (0, 2^{-N+1}] \cup \left[\sum_{n=1}^{N-1} (2^{-n}, 2^{-n-1}]\right] \cup \Lambda$$

while

$$(0,1) = \left(0, 2^{-N+1}\right) \cup \left[\sum_{n=1}^{N-1} \left(2^{-n}, 2^{-n-1}\right)\right] \cup \left\{2^{-n} : n = 1, 2, \dots, N\right\}$$

and so again it follows from what we have proved and Lemma 9.51 that $(0,1) \cong (0,1) \cup \Lambda$. Finally if $\Lambda = \{2, 3, 4, \dots\}$ is a countable set, we can show $(0,1) \cong (0,1) \cup \Lambda$ with the aid of the identities,

$$(0,1) = \left[\sum_{n=1}^\infty \left(2^{-n}, 2^{-n-1}\right)\right] \cup \left\{2^{-n} : n \in \mathbb{N}\right\}$$

and

$$(0,1) \cup \Lambda \cong (0,1] \cup \Lambda = \left[\sum_{n=1}^\infty (2^{-n}, 2^{-n-1}]\right] \cup \Lambda.$$

■

Notation 9.53 *Suppose* $(X,\mathcal{M})$ *is a measurable space and* A *is a set. Let* $\pi_a : X^A \to X$ *denote projection operator onto the* a^{th} *– component of* X^A *(i.e.* $\pi_a(\omega) = \omega(a)$ *for all* $a \in A$*) and let* $\mathcal{M}^{\otimes A} := \sigma(\pi_a : a \in A)$ *be the product* σ *– algebra on* X^A.

Lemma 9.54. *If* $\varphi : A \to B$ *is a bijection of sets and* $(X,\mathcal{M})$ *is a measurable space, then* $\left(X^A, \mathcal{M}^{\otimes A}\right) \cong \left(X^B, \mathcal{M}^{\otimes B}\right)$.

Proof. The map $f : X^B \to X^A$ defined by $f(\omega) = \omega \circ \varphi$ for all $\omega \in X^B$ is a bijection with $f^{-1}(\alpha) = \alpha \circ \varphi^{-1}$. If $a \in A$ and $\omega \in X^B$, we have

$$\pi_a^{X^A} \circ f(\omega) = f(\omega)(a) = \omega(\varphi(a)) = \pi_{\varphi(a)}^{X^B}(\omega),$$

where $\pi_a^{X^A}$ and $\pi_b^{X^B}$ are the projection operators on X^A and X^B respectively. Thus $\pi_a^{X^A} \circ f = \pi_{\varphi(a)}^{X^B}$ for all $a \in A$ which shows f is measurable. Similarly, $\pi_b^{X^B} \circ f^{-1} = \pi_{\varphi^{-1}(b)}^{X^A}$ showing f^{-1} is measurable as well. ■

Proposition 9.55. *Let* $\Omega := \{0,1\}^{\mathbb{N}}$, $\pi_i : \Omega \to \{0,1\}$ *be projection onto the* i^{th} *component, and* $\mathcal{B} := \sigma(\pi_1, \pi_2, \dots)$ *be the product* σ *– algebra on* Ω*. Then* $(\Omega, \mathcal{B}) \cong \left((0,1), \mathcal{B}_{(0,1)}\right)$.

Proof. We will begin by using a specific binary digit expansion of a point $x \in [0,1)$ to construct a map from $[0,1) \to \Omega$. To this end, let $r_1(x) = x$,

$$\gamma_1(x) := 1_{x \ge 2^{-1}} \text{ and } r_2(x) := x - 2^{-1}\gamma_1(x) \in (0, 2^{-1}),$$

then let $\gamma_2 := 1_{r_2 \ge 2^{-2}}$ and $r_3 = r_2 - 2^{-2}\gamma_2 \in \left(0, 2^{-2}\right)$. Working inductively, we construct $\{\gamma_k(x), r_k(x)\}_{k=1}^\infty$ such that $\gamma_k(x) \in \{0,1\}$, and

$$r_{k+1}(x) = r_k(x) - 2^{-k}\gamma_k(x) = x - \sum_{j=1}^k 2^{-j}\gamma_j(x) \in \left(0, 2^{-k}\right) \tag{9.69}$$

for all k. Let us now define $g : [0,1) \to \Omega$ by $g(x) := (\gamma_1(x), \gamma_2(x), \dots)$. Since each component function, $\pi_j \circ g = \gamma_j : [0,1) \to \{0,1\}$, is measurable it follows that g is measurable.

By construction,

$$x = \sum_{j=1}^k 2^{-j}\gamma_j(x) + r_{k+1}(x)$$

and $r_{k+1}(x) \to 0$ as $k \to \infty$, therefore

$$x = \sum_{j=1}^\infty 2^{-j}\gamma_j(x) \text{ and } r_{k+1}(x) = \sum_{j=k+1}^\infty 2^{-j}\gamma_j(x). \tag{9.70}$$

Hence if we define $f : \Omega \to [0,1]$ by $f = \sum_{j=1}^\infty 2^{-j}\pi_j$, then $f(g(x)) = x$ for all $x \in [0,1)$. This shows g is injective, f is surjective, and f in injective on the range of g.

We now claim that $\Omega_0 := g([0,1))$, the range of g, consists of those $\omega \in \Omega$ such that $\omega_i = 0$ for infinitely many i. Indeed, if there exists an $k \in \mathbb{N}$ such that $\gamma_j(x) = 1$ for all $j \ge k$, then (by Eq. (9.70)) $r_{k+1}(x) = 2^{-k}$ which would contradict Eq. (9.69). Hence $g([0,1)) \subset \Omega_0$. Conversely if $\omega \in \Omega_0$ and $x = f(\omega) \in [0,1)$, it is not hard to show inductively that $\gamma_j(x) = \omega_j$ for all j, i.e. $g(x) = \omega$. For example, if $\omega_1 = 1$ then $x \ge 2^{-1}$ and hence $\gamma_1(x) = 1$. Alternatively, if $\omega_1 = 0$, then

$$x = \sum_{j=2}^\infty 2^{-j}\omega_j < \sum_{j=2}^\infty 2^{-j} = 2^{-1}$$

so that $\gamma_1(x) = 0$. Hence it follows that $r_2(x) = \sum_{j=2}^\infty 2^{-j}\omega_j$ and by similar reasoning we learn $r_2(x) \ge 2^{-2}$ iff $\omega_2 = 1$, i.e. $\gamma_2(x) = 1$ iff $\omega_2 = 1$. The full induction argument is now left to the reader.

Since single point sets are in $\mathcal{B}$ and

$$\Lambda := \Omega \setminus \Omega_0 = \cup_{n=1}^\infty \{\omega \in \Omega : \omega_j = 1 \text{ for } j \ge n\}$$

is a countable set, it follows that $\Lambda \in \mathcal{B}$ and therefore $\Omega_0 = \Omega \setminus \Lambda \in \mathcal{B}$. Hence we may now conclude that $g : \left([0,1), \mathcal{B}_{[0,1)}\right) \to (\Omega_0, \mathcal{B}_{\Omega_0})$ is a measurable bijection with measurable inverse given by $f|_{\Omega_0}$, i.e. $\left([0,1), \mathcal{B}_{[0,1)}\right) \cong (\Omega_0, \mathcal{B}_{\Omega_0})$. An application of Lemma 9.51 and Proposition 9.52 now implies

$$\Omega = \Omega_0 \cup \Lambda \cong [0,1) \cup \mathbb{N} \cong [0,1) \cong (0,1).$$

■

Corollary 9.56. *The following spaces are all isomorphic to* $\left((0,1), \mathcal{B}_{(0,1)}\right)$*;* $(0,1)^d$ *and* $\mathbb{R}^d$ *for any* $d \in \mathbb{N}$ *and* $[0,1]^{\mathbb{N}}$ *and* $\mathbb{R}^{\mathbb{N}}$ *where both of these spaces are equipped with their natural product* σ *– algebras, .*

Proof. In light of Lemma 9.51 and Proposition 9.52 we know that $(0,1)^d \cong \mathbb{R}^d$ and $(0,1)^{\mathbb{N}} \cong [0,1]^{\mathbb{N}} \cong \mathbb{R}^{\mathbb{N}}$. So, using Proposition 9.55, it suffices to show $(0,1)^d \cong \Omega \cong (0,1)^{\mathbb{N}}$ and to do this it suffices to show $\Omega^d \cong \Omega$ and $\Omega^{\mathbb{N}} \cong \Omega$.

To reduce the problem further, let us observe that $\Omega^d \cong \{0,1\}^{\mathbb{N} \times \{1,2,\dots,d\}}$ and $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. For example, let $g : \Omega^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}^2}$ be defined by $g(\omega)(i,j) = \omega(i)(j)$ for all $\omega \in \Omega^{\mathbb{N}} = \left[\{0,1\}^{\mathbb{N}}\right]^{\mathbb{N}}$. Then g is a bijection and since $\pi_{(i,j)}^{\{0,1\}^{\mathbb{N}^2}} \circ g(\omega) = \pi_j^{\Omega}\left(\pi_i^{\Omega^{\mathbb{N}}}(\omega)\right)$, it follows that g is measurable. The inverse, $g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega^{\mathbb{N}}$, to g is given by $g^{-1}(\alpha)(i)(j) = \alpha(i,j)$. To see this map is measurable, we have $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} : \{0,1\}^{\mathbb{N}^2} \to \Omega = \{0,1\}^{\mathbb{N}}$ is given $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}(\alpha) = g^{-1}(\alpha)(i)(\cdot) = \alpha(i,\cdot)$ and hence

$$\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g(\alpha) = \alpha(i,j) = \pi_{i,j}^{\{0,1\}^{\mathbb{N}^2}}(\alpha)$$

from which it follows that $\pi_j^{\Omega} \circ \pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1} = \pi^{\{0,1\}^{\mathbb{N}^2}}$ is measurable for all $i, j \in \mathbb{N}$ and hence $\pi_i^{\Omega^{\mathbb{N}}} \circ g^{-1}$ is measurable for all $i \in \mathbb{N}$ and hence g^{-1} is measurable. This shows $\Omega^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}^2}$. The proof that $\Omega^d \cong \{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}}$ is analogous.

We may now complete the proof with a couple of applications of Lemma 9.54. Indeed $\mathbb{N}$, $\mathbb{N}\times\{1,2,\dots,d\}$, and $\mathbb{N}^2$ all have the same cardinality and therefore,

$$\{0,1\}^{\mathbb{N}\times\{1,2,\dots,d\}} \cong \{0,1\}^{\mathbb{N}^2} \cong \{0,1\}^{\mathbb{N}} = \Omega.$$

■

Corollary 9.57. *Suppose that $(X_n, \mathcal{M}_n)$ for $n \in \mathbb{N}$ are standard Borel spaces, then $X := \prod_{n=1}^{\infty} X_n$ equipped with the product σ – algebra, $\mathcal{M} := \otimes_{n=1}^{\infty}\mathcal{M}_n$ is again a standard Borel space.*

Proof. Let $A_n \in \mathcal{B}_{[0,1]}$ be Borel sets on $[0,1]$ such that there exists a measurable isomorpohism, $f_n : X_n \to A_n$. Then $f : X \to A := \prod_{n=1}^{\infty} A_n$ defined by $f(x_1, x_2, \dots) = (f_1(x_1), f_2(x_2), \dots)$ is easily seen to me a measure theoretic isomorphism when A is equipped with the product σ – algebra, $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$. So according to Corollary 9.56, to finish the proof it suffice to show $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n} = \mathcal{M}_A$ where $\mathcal{M} := \otimes_{n=1}^{\infty}\mathcal{B}_{[0,1]}$ is the product σ – algebra on $[0,1]^{\mathbb{N}}$.

The σ – algebra, $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$, is generated by sets of the form, $B := \prod_{n=1}^{\infty} B_n$ where $B_n \in \mathcal{B}_{A_n} \subset \mathcal{B}_{[0,1]}$. On the other hand, the σ – algebra, $\mathcal{M}_A$ is generated by sets of the form, $A \cap \tilde{B}$ where $\tilde{B} := \prod_{n=1}^{\infty} \tilde{B}_n$ with $\tilde{B}_n \in \mathcal{B}_{[0,1]}$. Since

$$A \cap \tilde{B} = \prod_{n=1}^{\infty}\left(\tilde{B}_n \cap A_n\right) = \prod_{n=1}^{\infty} B_n$$

where $B_n = \tilde{B}_n \cap A_n$ is the generic element in $\mathcal{B}_{A_n}$, we see that $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n}$ and $\mathcal{M}_A$ can both be generated by the same collections of sets, we may conclude that $\otimes_{n=1}^{\infty}\mathcal{B}_{A_n} = \mathcal{M}_A$. ■

Our next goal is to show that any Polish space with its Borel σ – algebra is a standard Borel space.

Notation 9.58 *Let $Q := [0,1]^{\mathbb{N}}$ denote the (infinite dimensional)* ***unit cube*** *in $\mathbb{R}^{\mathbb{N}}$. For $a, b \in Q$ let*

$$d(a,b) := \sum_{n=1}^{\infty} \frac{1}{2^n}|a_n - b_n| = \sum_{n=1}^{\infty} \frac{1}{2^n}|\pi_n(a) - \pi_n(b)|. \tag{9.71}$$

Exercise 9.11. Show d is a metric and that the Borel σ – algebra on (Q,d) is the same as the product σ – algebra.

Solution to Exercise (9.11). It is easily seen that d is a metric on Q which, by Eq. (9.71) is measurable relative to the product σ – algebra, $\mathcal{M}$.. Therefore, $\mathcal{M}$ contains all open balls and hence contains the Borel σ – algebra, $\mathcal{B}$. Conversely, since

$$|\pi_n(a) - \pi_n(b)| \le 2^n d(a,b),$$

each of the projection operators, $\pi_n : Q \to [0,1]$ is continuous. Therefore each π_n is $\mathcal{B}$ – measurable and hence $\mathcal{M} = \sigma(\{\pi_n\}_{n=1}^{\infty}) \subset \mathcal{B}$.

Theorem 9.59. *To every separable metric space (X,ρ), there exists a continuous injective map $G : X \to Q$ such that $G : X \to G(X) \subset Q$ is a homeomorphism. Moreover if the metric, ρ, is also complete, then $G(X)$ is a G_δ –set, i.e. the $G(X)$ is the countable intersection of open subsets of (Q,d). In short, any separable metrizable space X is homeomorphic to a subset of (Q,d) and if X is a Polish space then X is homeomorphic to a G_δ – subset of (Q,d).*

Proof. (This proof follows that in Rogers and Williams [11, Theorem 82.5 on p. 106.].) By replacing ρ by $\frac{\rho}{1+\rho}$ if necessary, we may assume that $0 \le \rho < 1$. Let $D = \{a_n\}_{n=1}^{\infty}$ be a countable dense subset of X and define

$$G(x) = (\rho(x,a_1), \rho(x,a_2), \rho(x,a_3), \dots) \in Q$$

and

$$\gamma(x,y) = d(G(x), G(y)) = \sum_{n=1}^{\infty} \frac{1}{2^n}|\rho(x,a_n) - \rho(y,a_n)|$$

for $x, y \in X$. To prove the first assertion, we must show G is injective and γ is a metric on X which is compatible with the topology determined by ρ.

If $G(x) = G(y)$, then $\rho(x,a) = \rho(y,a)$ for all $a \in D$. Since D is a dense subset of X, we may choose $\alpha_k \in D$ such that

$$0 = \lim_{k\to\infty}\rho(x,\alpha_k) = \lim_{k\to\infty}\rho(y,\alpha_k) = \rho(y,x)$$

and therefore $x = y$. A simple argument using the dominated convergence theorem shows $y \to \gamma(x,y)$ is ρ – continuous, i.e. $\gamma(x,y)$ is small if $\rho(x,y)$ is small. Conversely,

$$\begin{aligned}\rho(x,y) &\le \rho(x,a_n) + \rho(y,a_n) = 2\rho(x,a_n) + \rho(y,a_n) - \rho(x,a_n) \\ &\le 2\rho(x,a_n) + |\rho(x,a_n) - \rho(y,a_n)| \le 2\rho(x,a_n) + 2^n\gamma(x,y).\end{aligned}$$

Hence if $\varepsilon > 0$ is given, we may choose n so that $2\rho(x,a_n) < \varepsilon/2$ and so if $\gamma(x,y) < 2^{-(n+1)}\varepsilon$, it will follow that $\rho(x,y) < \varepsilon$. This shows $\tau_\gamma = \tau_\rho$. Since $G : (X,\gamma) \to (Q,d)$ is isometric, G is a homeomorphism.

Now suppose that (X,ρ) is a complete metric space. Let $S := G(X)$ and σ be the metric on S defined by $\sigma(G(x),G(y)) = \rho(x,y)$ for all $x,y \in X$. Then (S,σ) is a complete metric (being the isometric image of a complete metric space) and by what we have just prove, $\tau_\sigma = \tau_{d_S}$. Consequently, if $u \in S$ and $\varepsilon > 0$ is given, we may find $\delta'(\varepsilon)$ such that $B_\sigma(u,\delta'(\varepsilon)) \subset B_d(u,\varepsilon)$. Taking $\delta(\varepsilon) = \min(\delta'(\varepsilon),\varepsilon)$, we have $\operatorname{diam}_d(B_d(u,\delta(\varepsilon))) < \varepsilon$ and $\operatorname{diam}_\sigma(B_d(u,\delta(\varepsilon))) < \varepsilon$ where

$$\operatorname{diam}_\sigma(A) := \{\sup \sigma(u,v) : u,v \in A\} \text{ and}$$
$$\operatorname{diam}_d(A) := \{\sup d(u,v) : u,v \in A\}.$$

Let $\bar{S}$ denote the closure of S inside of (Q,d) and for each $n \in \mathbb{N}$ let

$$\mathcal{N}_n := \{N \in \tau_d : \operatorname{diam}_d(N) \vee \operatorname{diam}_\sigma(N \cap S) < 1/n\}$$

and let $U_n := \cup \mathcal{N}_n \in \tau_d$. From the previous paragraph, it follows that $S \subset U_n$ and therefore $S \subset \bar{S} \cap (\cap_{n=1}^\infty U_n)$.

Conversely if $u \in \bar{S} \cap (\cap_{n=1}^\infty U_n)$ and $n \in \mathbb{N}$, there exists $N_n \in \mathcal{N}_n$ such that $u \in N_n$. Moreover, since $N_1 \cap \cdots \cap N_n$ is an open neighborhood of $u \in \bar{S}$, there exists $u_n \in N_1 \cap \cdots \cap N_n \cap S$ for each $n \in \mathbb{N}$. From the definition of $\mathcal{N}_n$, we have $\lim_{n\to\infty} d(u,u_n) = 0$ and $\sigma(u_n,u_m) \le \max(n^{-1},m^{-1}) \to 0$ as $m,n \to \infty$. Since (S,σ) is complete, it follows that $\{u_n\}_{n=1}^\infty$ is convergent in (S,σ) to some element $u_0 \in S$. Since (S,d_S) has the same topology as (S,σ) it follows that $d(u_n,u_0) \to 0$ as well and thus that $u = u_0 \in S$. We have now shown, $S = \bar{S} \cap (\cap_{n=1}^\infty U_n)$. This completes the proof because we may write $\bar{S} = \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ where $S_{1/n} := \{u \in Q : d(u,\bar{S}) < 1/n\}$ and therefore, $S = \left(\bigcap_{n=1}^\infty U_n\right) \cap \left(\bigcap_{n=1}^\infty S_{1/n}\right)$ is a G_δ set. ■

Corollary 9.60. *Every Polish space, X, with its Borel σ – algebra is a standard Borel space. Consequently and Borel subset of X is also a standard Borel space.*

Proof. Theorem 9.59 shows that X is homeomorphic to a measurable (in fact a G_δ) subset Q_0 of (Q,d) and hence $X \cong Q_0$. Since Q is a standard Borel space so is Q_0 and hence so is X. ■

9.11 More Exercises

Exercise 9.12. 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 9.13. Prove the second assertion of Theorem 9.20. 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 9.14. (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 9.15. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 9.16. 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 9.17. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 9.18. 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 9.19. 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 9.20. Folland Problem 2.57 on p. 77.

Exercise 9.21. Folland Problem 2.58 on p. 77.

Exercise 9.22. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 9.23. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 9.24. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 9.25. Folland Problem 2.64 on p. 80. On the integrability of $|x|^a \left|\log|x|\right|^b$ for x near 0 and x near ∞ in $\mathbb{R}^n$.

Exercise 9.26. Show, using Problem 9.24 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

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^{d}\int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

Independence

As usual, $(\Omega, \mathcal{B}, P)$ will be some fixed probability space. Recall that for $A, B \in \mathcal{B}$ with $P(B) > 0$ we let

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

which is to be read as; **the probability of A given B.**

Definition 10.1. *We say that A is independent of B is $P(A|B) = P(A)$ or equivalently that*

$$P(A \cap B) = P(A) P(B).$$

We further say a finite sequence of collection of sets, $\{\mathcal{C}_i\}_{i=1}^n$, are independent if

$$P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$$

for all $A_i \in \mathcal{C}_i$ and $J \subset \{1, 2, \ldots, n\}$.

10.1 Basic Properties of Independence

If $\{\mathcal{C}_i\}_{i=1}^n$, are independent classes then so are $\{\mathcal{C}_i \cup \{\Omega\}\}_{i=1}^n$. Moreover, if we assume that $\Omega \in \mathcal{C}_i$ for each i, then $\{\mathcal{C}_i\}_{i=1}^n$, are independent iff

$$P\left(\cap_{j=1}^n A_j\right) = \prod_{j=1}^n P(A_j) \text{ for all } (A_1, \ldots, A_n) \in \mathcal{C}_1 \times \cdots \times \mathcal{C}_n.$$

Theorem 10.2. *Suppose that $\{\mathcal{C}_i\}_{i=1}^n$ is a finite sequence of independent π – classes. Then $\{\sigma(\mathcal{C}_i)\}_{i=1}^n$ are also independent.*

Proof. As mentioned above, we may always assume without loss of generality that $\Omega \in \mathcal{C}_i$. Fix, $A_j \in \mathcal{C}_j$ for $j = 2, 3, \ldots, n$. We will begin by showing that

$$Q(A) := P(A \cap A_2 \cap \cdots \cap A_n) = P(A) P(A_2) \ldots P(A_n) \text{ for all } A \in \sigma(\mathcal{C}_1). \tag{10.1}$$

Since $Q(\cdot)$ and $P(A_2) \ldots P(A_n) P(\cdot)$ are both finite measures agreeing on Ω and A in the π – system $\mathcal{C}_1$, Eq. (10.1) is a direct consequence of Proposition 5.15. Since $(A_2, \ldots, A_n) \in \mathcal{C}_2 \times \cdots \times \mathcal{C}_n$ were arbitrary we may now conclude that $\sigma(\mathcal{C}_1), \mathcal{C}_2, \ldots, \mathcal{C}_n$ are independent.

By applying the result we have just proved to the sequence, $\mathcal{C}_2, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ shows that $\sigma(\mathcal{C}_2), \mathcal{C}_3, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ are independent. Similarly we show inductively that

$$\sigma(\mathcal{C}_j), \mathcal{C}_{j+1}, \ldots, \mathcal{C}_n, \sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_{j-1})$$

are independent for each $j = 1, 2, \ldots, n$. The desired result occurs at $j = n$. ■

Definition 10.3. *Let $(\Omega, , \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. The maps $\{Y_i\}_{i=1}^n$ are P - independent iff $\{\mathcal{C}_i\}_{i=1}^n$ are P – independent, where $\mathcal{C}_i := Y_i^{-1}(\mathcal{F}_i) = \sigma(Y_i) \subset \mathcal{B}$ for $1 \le i \le n$.*

Theorem 10.4 (Independence and Product Measures). *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\{(S_i, \mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i : \Omega \to S_i$ be a measurable map for $1 \le i \le n$. Further let $\mu_i := P \circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$. Then $\{Y_i\}_{i=1}^n$ are independent iff*

$$\mathrm{Law}_P(Y_1, \ldots, Y_n) = \mu_1 \otimes \cdots \otimes \mu_n,$$

where $(Y_1, \ldots, Y_n) : \Omega \to S_1 \times \cdots \times S_n$ and

$$\mathrm{Law}_P(Y_1, \ldots, Y_n) = P \circ (Y_1, \ldots, Y_n)^{-1} : \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n \to [0, 1]$$

is the joint law of $Y_1, \ldots, Y_n$.

Proof. Recall that the general element of $\mathcal{C}_i$ is of the form $A_i = Y_i^{-1}(B_i)$ with $B_i \in \mathcal{S}_i$. Therefore for $A_i = Y_i^{-1}(B_i) \in \mathcal{C}_i$ we have

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P((Y_1, \ldots, Y_n) \in B_1 \times \cdots \times B_n) \\ &= ((Y_1, \ldots, Y_n)_* P)(B_1 \times \cdots \times B_n). \end{aligned}$$

If $(Y_1, \ldots, Y_n)_* P = \mu_1 \otimes \cdots \otimes \mu_n$ it follows that

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= \mu_1 \otimes \cdots \otimes \mu_n (B_1 \times \cdots \times B_n) \\ &= \mu_1(B_1) \cdots \mu(B_n) = P(Y_1 \in B_1) \cdots P(Y_n \in B_n) \\ &= P(A_1) \ldots P(A_n) \end{aligned}$$

and therefore $\{\mathcal{C}_i\}$ are P – independent and hence $\{Y_i\}$ are P – independent.

Conversely if $\{Y_i\}$ are P – independent, i.e. $\{\mathcal{C}_i\}$ are P – independent, then

$$\begin{aligned}P((Y_1,\dots,Y_n)\in B_1\times\dots\times B_n) &= P(A_1\cap\dots\cap A_n)\\ &= P(A_1)\dots P(A_n)\\ &= P(Y_1\in B_1)\cdots P(Y_n\in B_n)\\ &= \mu_1(B_1)\cdots\mu(B_n)\\ &= \mu_1\otimes\dots\otimes\mu_n(B_1\times\dots\times B_n).\end{aligned}$$

Since

$$\pi := \{B_1\times\dots\times B_n : B_i\in\mathcal{S}_i \text{ for } 1\le i\le n\}$$

is a π – system which generates $\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$ and

$$(Y_1,\dots,Y_n)_* P = \mu_1\otimes\dots\otimes\mu_n \text{ on } \pi,$$

it follows that $(Y_1,\dots,Y_n)_* P = \mu_1\otimes\dots\otimes\mu_n$ on all of $\mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$. ■

Remark 10.5. When have a collection of not necessarily independent random functions, $Y_i : \Omega\to S_i$, like in Theorem 10.4 it is **not** in general possible to recover the joint distribution, $\pi := \mathrm{Law}_P(Y_1,\dots,Y_n)$, from the individual distributions, $\mu_i = \mathrm{Law}_P(Y_i)$ for all $1\le i\le n$. For example suppose that $S_i=\mathbb{R}$ for $i=1,2$. μ is a probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$, and (Y_1,Y_2) have joint distribution, π, given by,

$$\pi(C) = \int_{\mathbb{R}} 1_C(x,x)\,d\mu(x) \text{ for all } C\in\mathcal{B}_{\mathbb{R}}.$$

If we let $\mu_i = \mathrm{Law}(Y_i)$, then for all $A\in\mathcal{B}_{\mathbb{R}}$ we have

$$\begin{aligned}\mu_1(A) &= P(Y_1\in A) = P((Y_1,Y_2)\in A\times\mathbb{R})\\ &= \pi(A\times\mathbb{R}) = \int_{\mathbb{R}} 1_{A\times\mathbb{R}}(x,x)\,d\mu(x) = \mu(A).\end{aligned}$$

Similarly we show that $\mu_2=\mu$. On the other hand if μ is not concentrated on one point, $\mu\otimes\mu$ is another probability measure on $(\mathbb{R}^2,\mathcal{B}_{\mathbb{R}^2})$ with the same **marginals** as π, i.e. $\pi(A\times\mathbb{R}) = \mu(A) = \pi(\mathbb{R}\times A)$ for all $A\in\mathcal{B}_{\mathbb{R}}$.

Lemma 10.6. *Let $(\Omega,,\mathcal{B},P)$ be a probability space, $\{(S_i,\mathcal{S}_i)\}_{i=1}^n$ and $\{(T_i,\mathcal{T}_i)\}_{i=1}^n$ be two collection of measurable spaces, $F_i : S_i\to T_i$ be a measurable map for each i and $Y_i:\Omega\to S_i$ be a collection of P – independent measurable maps. Then $\{F_i\circ Y_i\}_{i=1}^n$ are also P – independent.*

Proof. Notice that

$$\sigma(F_i\circ Y_i) = (F_i\circ Y_i)^{-1}(\mathcal{T}_i) = Y_i^{-1}\left(F_i^{-1}(\mathcal{T}_i)\right)\subset Y_i^{-1}(\mathcal{S}_i) = \mathcal{C}_i.$$

The fact that $\{\sigma(F_i\circ Y_i)\}_{i=1}^n$ is independent now follows easily from the assumption that $\{\mathcal{C}_i\}$ are P – independent. ■

Example 10.7. If $\Omega := \prod_{i=1}^n S_i$, $\mathcal{B} := \mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$, $Y_i(s_1,\dots,s_n) = s_i$ for all $(s_1,\dots,s_n)\in\Omega$, and $\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i)$ for all i. Then the probability measures, P, on $(\Omega,\mathcal{B})$ for which $\{\mathcal{C}_i\}_{i=1}^n$ are independent are precisely the product measures, $P=\mu_1\otimes\dots\otimes\mu_n$ where μ_i is a probability measure on $(S_i,\mathcal{S}_i)$ for $1\le i\le n$. Notice that in this setting,

$$\mathcal{C}_i := Y_i^{-1}(\mathcal{S}_i) = \{S_1\times\dots\times S_{i-1}\times B\times S_{i+1}\times\dots\times S_n : B\in\mathcal{S}_i\}\subset\mathcal{B}.$$

Proposition 10.8. *Suppose that $(\Omega,\mathcal{B},P)$ is a probability space and $\{Z_j\}_{j=1}^n$ are independent integrable random variables. Then $\prod_{j=1}^n Z_j$ is also integrable and*

$$\mathbb{E}\left[\prod_{j=1}^n Z_j\right] = \prod_{j=1}^n \mathbb{E}Z_j.$$

Proof. Let $\mu_j := P\circ Z_j^{-1} : \mathcal{B}_{\mathbb{R}}\to[0,1]$ be the law of Z_j for each j. Then we know $(Z_1,\dots,Z_n)_* P = \mu_1\otimes\dots\otimes\mu_n$. Therefore by Example 7.52 and Tonelli's theorem,

$$\begin{aligned}\mathbb{E}\left[\prod_{j=1}^n |Z_j|\right] &= \int_{\mathbb{R}^n}\left[\prod_{j=1}^n |z_j|\right] d\left(\otimes_{j=1}^n\mu_j\right)(z)\\ &= \prod_{j=1}^n\int_{\mathbb{R}^n}|z_j|\,d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E}|Z_j| < \infty\end{aligned}$$

which shows that $\prod_{j=1}^n Z_j$ is integrable. Thus again by Example 7.52 and Fubini's theorem,

$$\begin{aligned}\mathbb{E}\left[\prod_{j=1}^n Z_j\right] &= \int_{\mathbb{R}^n}\left[\prod_{j=1}^n z_j\right] d\left(\otimes_{j=1}^n\mu_j\right)(z)\\ &= \prod_{j=1}^n\int_{\mathbb{R}} z_j\,d\mu_j(z_j) = \prod_{j=1}^n \mathbb{E}Z_j.\end{aligned}$$

■

Theorem 10.9. *Let $(\Omega,,\mathcal{B},P)$ be a probability space, $\{(S_i,\mathcal{S}_i)\}_{i=1}^n$ be a collection of measurable spaces and $Y_i:\Omega\to S_i$ be a measurable map for $1\le i\le n$. Further let $\mu_i := P\circ Y_i^{-1} = \mathrm{Law}_P(Y_i)$ and $\pi := P\circ(Y_1,\dots,Y_n)^{-1} : \mathcal{S}_1\otimes\dots\otimes\mathcal{S}_n$ be the joint distribution of*

$$(Y_1,\dots,Y_n):\Omega\to S_1\times\dots\times S_n.$$

Then the following are equivalent,

1. $\{Y_i\}_{i=1}^n$ *are independent,*
2. $\pi = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$
3. *for all bounded measurable functions,* $f : (S_1 \times \cdots \times S_n, \mathcal{S}_1 \otimes \cdots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,

$$\mathbb{E}f(Y_1, \dots, Y_n) = \int_{S_1 \times \cdots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \quad (10.2)$$

(where the integrals may be taken in any order),

4. $\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(Y_i)]$ *for all bounded (or non-negative) measurable functions,* $f_i : S_i \to \mathbb{R}$ *or* $\mathbb{C}$.

Proof. ($1 \iff 2$) has already been proved in Theorem 10.4. The fact that ($2. \implies 3.$) now follows from Exercise 7.11 and Fubini's theorem. Similarly, ($3. \implies 4.$) follows from Exercise 7.11 and Fubini's theorem after taking $f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$. Lastly for ($4. \implies 1.$), let $A_i \in \mathcal{S}_i$ and take $f_i := 1_{A_i}$ in 4. to learn,

$$P\left(\cap_{i=1}^n \{Y_i \in A_i\}\right) = \mathbb{E}\left[\prod_{i=1}^n 1_{A_i}(Y_i)\right] = \prod_{i=1}^n \mathbb{E}[1_{A_i}(Y_i)] = \prod_{i=1}^n P(Y_i \in A_i)$$

which shows that the $\{Y_i\}_{i=1}^n$ are independent. ■

Corollary 10.10. *Suppose that* $(\Omega, \mathcal{B}, P)$ *is a probability space and* $\{Y_j : \Omega \to \mathbb{R}\}_{j=1}^n$ *is a sequence of random variables with countable ranges, say* $\Lambda \subset \mathbb{R}$. *Then* $\{Y_j\}_{j=1}^n$ *are independent iff*

$$P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) = \prod_{j=1}^n P(Y_j = y_j) \quad (10.3)$$

for all choices of $y_1, \dots, y_n \in \Lambda$.

Proof. If the $\{Y_j\}$ are independent then clearly Eq. (10.3) holds by definition as $\{Y_j = y_j\} \in Y_j^{-1}(\mathcal{B}_{\mathbb{R}})$. Conversely if Eq. (10.3) holds and $f_i : \mathbb{R} \to [0, \infty)$ are measurable functions then,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^n f_i(Y_i)\right] &= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot P\left(\cap_{j=1}^n \{Y_j = y_j\}\right) \\
&= \sum_{y_1, \dots, y_n \in \Lambda} \prod_{i=1}^n f_i(y_i) \cdot \prod_{j=1}^n P(Y_j = y_j) \\
&= \prod_{i=1}^n \sum_{y_i \in \Lambda} f_i(y_i) \cdot P(Y_j = y_j) \\
&= \prod_{i=1}^n \mathbb{E}[f_i(Y_i)]
\end{aligned}$$

wherein we have used Tonelli's theorem for sum in the third equality. It now follows that $\{Y_i\}$ are independent using item 4. of Theorem 10.9. ■

Exercise 10.1. Suppose that $\Omega = (0, 1]$, $\mathcal{B} = \mathcal{B}_{(0,1]}$, and $P = m$ is Lebesgue measure on $\mathcal{B}$. Let $Y_i(\omega) := \omega_i$ be the i^{th} – digit in the base two expansion of ω. To be more precise, the $Y_i(\omega) \in \{0, 1\}$ is chosen so that

$$\omega = \sum_{i=1}^\infty Y_i(\omega) 2^{-i} \text{ for all } \omega_i \in \{0, 1\}.$$

As long as $\omega \neq k2^{-n}$ for some $0 < k \leq n$, the above equation uniquely determines the $\{Y_i(\omega)\}$. Owing to the fact that $\sum_{l=n+1}^\infty 2^{-l} = 2^{-n}$, if $\omega = k2^{-n}$, there is some ambiguity in the definitions of the $Y_i(\omega)$ for large i which you may resolve anyway you choose. Show the random variables, $\{Y_i\}_{i=1}^n$, are i.i.d. for each $n \in \mathbb{N}$ with $P(Y_i = 1) = 1/2 = P(Y_i = 0)$ for all i.

Hint: the idea is that knowledge of $(Y_1(\omega), \dots, Y_n(\omega))$ is equivalent to knowing for which $k \in \mathbb{N}_0 \cap [0, 2^n)$ that $\omega \in (2^{-n}k, 2^{-n}(k+1)]$ and that this knowledge in no way helps you predict the value of $Y_{n+1}(\omega)$. More formally, you might start by showing,

$$P\left(\{Y_{n+1} = 1\} | (2^{-n}k, 2^{-n}(k+1)]\right) = \frac{1}{2} = P\left(\{Y_{n+1} = 0\} | (2^{-n}k, 2^{-n}(k+1)]\right).$$

See Section 10.9 if you need some more help with this exercise.

Exercise 10.2. Let X, Y be two random variables on $(\Omega, \mathcal{B}, P)$.

1. Show that X and Y are independent iff $\mathrm{Cov}(f(X), g(Y)) = 0$ (i.e. $f(X)$ and $g(Y)$ are **uncorrelated**) for bounded measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$.
2. If $X, Y \in L^2(P)$ and X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.
3. Show by example that if $X, Y \in L^2(P)$ and $\mathrm{Cov}(X, Y) = 0$ does not necessarily imply that X and Y are independent. **Hint:** try taking $(X, Y) = (X, ZX)$ where X and Z are independent simple random variables such that $\mathbb{E}Z = 0$ similar to Remark 9.40.

Solution to Exercise (10.2). 1. Since

$$\mathrm{Cov}(f(X), g(Y)) = \mathbb{E}[f(X) g(Y)] - \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)]$$

it follows that $\mathrm{Cov}(f(X), g(Y)) = 0$ iff

$$\mathbb{E}[f(X) g(Y)] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)]$$

from which item 1. easily follows.

2. Let $f_M(x) = x 1_{|x| \leq M}$, then by independence,

$$\mathbb{E}\left[f_M(X) f_M(Y)\right] = \mathbb{E}\left[f_M(X)\right] \mathbb{E}\left[f_M(Y)\right]. \tag{10.4}$$

Since

$$|f_M(X) f_M(Y)| \le |XY| \le \frac{1}{2}\left(X^2 + Y^2\right) \in L^1(P),$$
$$|f_M(X)| \le |X| \le \frac{1}{2}\left(1 + X^2\right) \in L^1(P), \text{ and}$$
$$|f_M(Y)| \le |Y| \le \frac{1}{2}\left(1 + Y^2\right) \in L^1(P),$$

we may use the DCT three times to pass to the limit as $M \to \infty$ in Eq. (10.4) to learn that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, i.e. $\mathrm{Cov}(X,Y) = 0$.

3. Let X and Z be independent with $P(Z = \pm 1) = \frac{1}{2}$ and take $Y = XZ$. Then $\mathbb{E}Z = 0$ and

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}\left[X^2 Z\right] - \mathbb{E}[X]\mathbb{E}[XZ] \\ &= \mathbb{E}\left[X^2\right] \cdot \mathbb{E}Z - \mathbb{E}[X]\mathbb{E}[X]\mathbb{E}Z = 0.\end{aligned}$$

On the other hand it should be intuitively clear that X and Y are not independent since knowledge of X typically will give some information about Y. To verify this assertion let us suppose that X is a discrete random variable with $P(X = 0) = 0$. Then

$$P(X = x, Y = y) = P(X = x, xZ = y) = P(X = x) \cdot P(X = y/x)$$

while

$$P(X = x) P(Y = y) = P(X = x) \cdot P(XZ = y).$$

Thus for X and Y to be independent we would have to have,

$$P(xX = y) = P(XZ = y) \text{ for all } x, y.$$

This is clearly not going to be true in general. For example, suppose that $P(X = 1) = \frac{1}{2} = P(X = 0)$. Taking $x = y = 1$ in the previously displayed equation would imply

$$\frac{1}{2} = P(X = 1) = P(XZ = 1) = P(X = 1, Z = 1) = P(X = 1) P(Z = 1) = \frac{1}{4}$$

which is false.

Let us now specialize to the case where $S_i = \mathbb{R}^{m_i}$ and $\mathcal{S}_i = \mathcal{B}_{\mathbb{R}^{m_i}}$ for some $m_i \in \mathbb{N}$.

Theorem 10.11. *Let $(\Omega, \mathcal{B}, P)$ be a probability space, $m_j \in \mathbb{N}$, $S_j = \mathbb{R}^{m_j}$, $\mathcal{S}_j = \mathcal{B}_{\mathbb{R}^{m_j}}$, $Y_j : \Omega \to S_j$ be random vectors, and $\mu_j := \mathrm{Law}_P(Y_j) = P \circ Y_j^{-1} : \mathcal{S}_j \to [0,1]$ for $1 \le j \le n$. The the following are equivalent;*

1. *$\{Y_j\}_{j=1}^n$ are independent,*
2. *$\mathrm{Law}_P(Y_1, \dots, Y_n) = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$*
3. *for all bounded measurable functions, $f : (S_1 \times \dots \times S_n, \mathcal{S}_1 \otimes \dots \otimes \mathcal{S}_n) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,*
$$\mathbb{E} f(Y_1, \dots, Y_n) = \int_{S_1 \times \dots \times S_n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \tag{10.5}$$
(where the integrals may be taken in any order),
4. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[f_j(Y_j)\right]$ for all bounded (or non-negative) measurable functions, $f_j : S_j \to \mathbb{R}$ or $\mathbb{C}$.*
5. *$P\left(\cap_{j=1}^n \{Y_j \le y_j\}\right) = \prod_{j=1}^n P(\{Y_j \le y_j\})$ for all $y_j \in S_j$, where we say that $Y_j \le y_j$ iff $(Y_j)_k \le (y_j)_k$ for $1 \le k \le m_j$.*
6. *$\mathbb{E}\left[\prod_{j=1}^n f_j(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[f_j(Y_j)\right]$ for all $f_j \in C_c(S_j, \mathbb{R})$,*
7. *$\mathbb{E}\left[e^{i \sum_{j=1}^n \lambda_j \cdot Y_j}\right] = \prod_{j=1}^n \mathbb{E}\left[e^{i\lambda_j \cdot Y_j}\right]$ for all $\lambda_j \in S_j = \mathbb{R}^{m_j}$.*

Proof. The equivalence of 1. – 4. has already been proved in Theorem 10.9. It is also clear that item 4. implies both or items 5. –7. upon noting that item 5. may be written as,

$$\mathbb{E}\left[\prod_{j=1}^n 1_{(-\infty, y_j]}(Y_j)\right] = \prod_{j=1}^n \mathbb{E}\left[1_{(-\infty, y_j]}(Y_j)\right]$$

where $(-\infty, y_j] := (-\infty, (y_j)_1] \times \dots \times (-\infty, (y_j)_{m_j}]$. The proofs that either 5. or 6. or 7. implies item 3. is a simple application of the multiplicative system theorem in the form of either Corollary 8.3 or Corollary 8.8. In each case, let $\mathbb{H}$ denote the linear space of bounded measurable functions such that Eq. (10.5) holds. To complete the proof I will simply give you the multiplicative system, $\mathbb{M}$, to use in each of the cases. To describe $\mathbb{M}$, let $N = m_1 + \dots + m_n$ and

$$y = (y_1, \dots, y_n) = \left(y^1, y^2, \dots, y^N\right) \in \mathbb{R}^N \text{ and}$$
$$\lambda = (\lambda_1, \dots, \lambda_n) = \left(\lambda^1, \lambda^2, \dots, \lambda^N\right) \in \mathbb{R}^N$$

For showing 5. $\implies$ 3.take $\mathbb{M} = \left\{1_{(-\infty, y]} : y \in \mathbb{R}^N\right\}$.

For showing 6. $\implies$ 3. take $\mathbb{M}$ to be a those functions on $\mathbb{R}^N$ which are of the form, $f(y) = \prod_{l=1}^N f_l\left(y^l\right)$ with each $f_l \in C_c(\mathbb{R})$.

For showing 7. $\implies$ 3. take $\mathbb{M}$ to be the functions of the form,

$$f(y) = \exp\left(i \sum_{j=1}^n \lambda_j \cdot y_j\right) = \exp(i\lambda \cdot y).$$

■

Definition 10.12. *A collection of subsets of $\mathcal{B}$, $\{\mathcal{C}_t\}_{t\in T}$ is said to be independent iff $\{\mathcal{C}_t\}_{t\in\Lambda}$ are independent for all finite subsets, $\Lambda \subset T$. More explicitly, we are requiring*

$$P\left(\cap_{t\in\Lambda} A_t\right) = \prod_{t\in\Lambda} P\left(A_t\right)$$

whenever Λ is a finite subset of T and $A_t \in \mathcal{C}_t$ for all $t \in \Lambda$.

Corollary 10.13. *If $\{\mathcal{C}_t\}_{t\in T}$ is a collection of independent classes such that each $\mathcal{C}_t$ is a π – system, then $\{\sigma(\mathcal{C}_t)\}_{t\in T}$ are independent as well.*

Definition 10.14. *A collections of random variables, $\{X_t : t \in T\}$ are* ***independent*** *iff $\{\sigma(X_t) : t \in T\}$ are independent.*

Example 10.15. Suppose that $\{\mu_n\}_{n=1}^\infty$ is any sequence of probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Let $\Omega = \mathbb{R}^\mathbb{N}$, $\mathcal{B} := \otimes_{n=1}^\infty \mathcal{B}_\mathbb{R}$ be the product σ – algebra on Ω, and $P := \otimes_{n=1}^\infty \mu_n$ be the product measure. Then the random variables, $\{Y_n\}_{n=1}^\infty$ defined by $Y_n(\omega) = \omega_n$ for all $\omega \in \Omega$ are independent with $\mathrm{Law}_P(Y_n) = \mu_n$ for each n.

Lemma 10.16 (Independence of groupings). *Suppose that $\{\mathcal{B}_t : t \in T\}$ is an independent family of σ – fields. Suppose further that $\{T_s\}_{s\in S}$ is a partition of T (i.e. $T = \sum_{s\in S} T_s$) and let*

$$\mathcal{B}_{T_s} = \vee_{t\in T_s}\mathcal{B}_t = \sigma\left(\cup_{t\in T_s}\mathcal{B}_t\right).$$

Then $\{\mathcal{B}_{T_s}\}_{s\in S}$ is again independent family of σ fields.

Proof. Let

$$\mathcal{C}_s = \left\{\cap_{\alpha\in K} B_\alpha : B_\alpha \in \mathcal{B}_\alpha,\ K \subset\subset T_s\right\}.$$

It is now easily checked that $\mathcal{B}_{T_s} = \sigma(\mathcal{C}_s)$ and that $\{\mathcal{C}_s\}_{s\in S}$ is an independent family of π – systems. Therefore $\{\mathcal{B}_{T_s}\}_{s\in S}$ is an independent family of σ – algebras by Corollary 10.13. ∎

Corollary 10.17. *Suppose that $\{Y_n\}_{n=1}^\infty$ is a sequence of independent random variables (or vectors) and $\Lambda_1, \dots, \Lambda_m$ is a collection of pairwise disjoint subsets of $\mathbb{N}$. Further suppose that $f_i : \mathbb{R}^{\Lambda_i} \to \mathbb{R}$ is a measurable function for each $1 \le i \le m$, then $Z_i := f_i\left(\{Y_l\}_{l\in\Lambda_i}\right)$ is again a collection of independent random variables.*

Proof. Notice that $\sigma(Z_i) \subset \sigma\left(\{Y_l\}_{l\in\Lambda_i}\right) = \sigma\left(\cup_{l\in\Lambda_i}\sigma(Y_l)\right)$. Since $\{\sigma(Y_l)\}_{l=1}^\infty$ are independent by assumption, it follows from Lemma 10.16 that $\left\{\sigma\left(\{Y_l\}_{l\in\Lambda_i}\right)\right\}_{i=1}^m$ are independent and therefore so is $\{\sigma(Z_i)\}_{i=1}^m$, i.e. $\{Z_i\}_{i=1}^m$ are independent. ∎

Definition 10.18 (i.i.d.). *A sequences of random variables, $\{X_n\}_{n=1}^\infty$, on a probability space, $(\Omega, \mathcal{B}, P)$, are* ***i.i.d. (= independent and identically distributed)*** *if they are independent and $(X_n)_* P = (X_k)_* P$ for all k, n. That is we should have*

$$P(X_n \in A) = P(X_k \in A) \text{ for all } k, n \in \mathbb{N} \text{ and } A \in \mathcal{B}_\mathbb{R}.$$

Observe that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables iff

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{j=1}^n P(X_i \in A_i) = \prod_{j=1}^n P(X_1 \in A_i) = \prod_{j=1}^n \mu(A_i) \tag{10.6}$$

where $\mu = (X_1)_* P$. The identity in Eq. (10.6) is to hold for all $n \in \mathbb{N}$ and all $A_i \in \mathcal{B}_\mathbb{R}$. If we choose $\mu_n = \mu$ in Example 10.15, the $\{Y_n\}_{n=1}^\infty$ there are i.i.d. with $\mathrm{Law}_P(Y_n) = P \circ Y_n^{-1} = \mu$ for all $n \in \mathbb{N}$.

The following theorem follows immediately from the definitions and Theorem 10.11.

Theorem 10.19. *Let $\mathbb{X} := \{X_t : t \in T\}$ be a collection of random variables. Then the following are equivalent:*

1. *The collection $\mathbb{X}$ is independent,*
2. $$P\left(\cap_{t\in\Lambda}\{X_t \in A_t\}\right) = \prod_{t\in\Lambda} P(X_t \in A_t)$$
   for all finite subsets, $\Lambda \subset T$, and all $\{A_t\}_{t\in\Lambda} \subset \mathcal{B}_\mathbb{R}$.
3. $$P\left(\cap_{t\in\Lambda}\{X_t \le x_t\}\right) = \prod_{t\in\Lambda} P(X_t \le x_t)$$
   for all finite subsets, $\Lambda \subset T$, and all $\{x_t\}_{t\in\Lambda} \subset \mathbb{R}$.
4. *For all $\Gamma \subset\subset T$ and $f_t : \mathbb{R}^n \to \mathbb{R}$ which are bounded an measurable for all $t \in \Gamma$,*
   $$\mathbb{E}\left[\prod_{t\in\Gamma} f_t(X_t)\right] = \prod_{t\in\Gamma}\mathbb{E} f_t(X_t) = \int_{\mathbb{R}^\Gamma}\prod_{t\in\Gamma} f_t(x_t)\prod_{t\in\Gamma} d\mu_t(x_t).$$
5. $\mathbb{E}\left[\prod_{t\in\Gamma}\exp\left(e^{i\lambda_t\cdot X_t}\right)\right] = \prod_{t\in\Gamma}\hat{\mu}_t(\lambda)$.
6. *For all $\Gamma \subset\subset T$ and $f : (\mathbb{R}^n)^\Gamma \to \mathbb{R}$,*
   $$\mathbb{E}\left[f(X_\Gamma)\right] = \int_{(\mathbb{R}^n)^\Gamma} f(x)\prod_{t\in\Gamma} d\mu_t(x_t).$$
7. *For all $\Gamma \subset\subset T$, $\mathrm{Law}_P(X_\Gamma) = \otimes_{t\in\Gamma}\mu_t$.*

8. $\mathrm{Law}_P(X) = \otimes_{t\in T}\mu_t$.

Moreover, if $\mathcal{B}_t$ is a sub-σ - algebra of $\mathcal{B}$ for $t \in T$, then $\{\mathcal{B}_t\}_{t\in T}$ are independent iff for all $\Gamma \subset\subset T$,

$$\mathbb{E}\left[\prod_{t\in\Gamma} X_t\right] = \prod_{t\in\Gamma}\mathbb{E}X_t \text{ for all } X_t \in L^\infty(\Omega, \mathcal{B}_t, P).$$

Proof. The equivalence of 1. and 2. follows almost immediately form the definition of independence and the fact that $\sigma(X_t) = \{\{X_t \in A\} : A \in \mathcal{B}_\mathbb{R}\}$. Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 10.13 with $\mathcal{C}_t := \{\{X_t \le a\} : a \in \mathbb{R}\}$ and making use the observations that $\mathcal{C}_t$ is a π – system for all t and that $\sigma(\mathcal{C}_t) = \sigma(X_t)$. The remaining equivalence are also easy to check. ■

10.2 Examples of Independence

10.2.1 An Example of Ranks

Lemma 10.20 (No Ties). *Suppose that X and Y are independent random variables on a probability space $(\Omega, \mathcal{B}, P)$. If $F(x) := P(X \le x)$ is continuous, then $P(X = Y) = 0$.*

Proof. Let $\mu(A) := P(X \in A)$ and $\nu(A) = P(Y \in A)$. Because F is continuous, $\mu(\{y\}) = F(y) - F(y-) = 0$, and hence

$$\begin{aligned}P(X = Y) &= \mathbb{E}\left[1_{\{X=Y\}}\right] = \int_{\mathbb{R}^2} 1_{\{x=y\}} d(\mu\otimes\nu)(x,y)\\ &= \int_\mathbb{R} d\nu(y)\int_\mathbb{R} d\mu(x)\, 1_{\{x=y\}} = \int_\mathbb{R}\mu(\{y\})\, d\nu(y)\\ &= \int_\mathbb{R} 0\ d\nu(y) = 0.\end{aligned}$$

Second Proof. For sake of comparison, lets give a proof where we do not allow ourselves to use Fubini's theorem. To this end let $\left\{a_l := \frac{l}{N}\right\}_{l=-\infty}^\infty$ (or for the moment any sequence such that, $a_l < a_{l+1}$ for all $l \in \mathbb{Z}$, $\lim_{l\to\pm\infty} a_l = \pm\infty$). Then

$$\{(x,x) : x \in \mathbb{R}\} \subset \cup_{l\in\mathbb{Z}}\left[(a_l, a_{l+1}] \times (a_l, a_{l+1}]\right]$$

and therefore,

$$\begin{aligned}P(X = Y) &\le \sum_{l\in\mathbb{Z}} P(X \in (a_l, a_{l+1}], Y \in (a_l, a_{l+1}]) = \sum_{l\in\mathbb{Z}}\left[F(a_{l+1}) - F(a_l)\right]^2\\ &\le \sup_{l\in\mathbb{Z}}\left[F(a_{l+1}) - F(a_l)\right]\sum_{l\in\mathbb{Z}}\left[F(a_{l+1}) - F(a_l)\right] = \sup_{l\in\mathbb{Z}}\left[F(a_{l+1}) - F(a_l)\right].\end{aligned}$$

Since F is continuous and $F(\infty+) = 1$ and $F(\infty-) = 0$, it is easily seen that F is uniformly continuous on $\mathbb{R}$. Therefore, if we choose $a_l = \frac{l}{N}$, we have

$$P(X = Y) \le \limsup_{N\to\infty}\sup_{l\in\mathbb{Z}}\left[F\left(\frac{l+1}{N}\right) - F\left(\frac{l}{N}\right)\right] = 0.$$

■

Let $\{X_n\}_{n=1}^\infty$ be i.i.d. with common continuous distribution function, F. So by Lemma 10.20 we know that

$$P(X_i = X_j) = 0 \text{ for all } i \ne j.$$

Let R_n denote the "**rank**" of X_n in the list $(X_1, \dots, X_n)$, i.e.

$$R_n := \sum_{j=1}^n 1_{X_j \ge X_n} = \#\{j \le n : X_j \ge X_n\}.$$

Thus $R_n = k$ if X_n is the k^{th} – largest element in the list, $(X_1, \dots, X_n)$. For example if $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, \dots) = (9, -8, 3, 7, 23, 0, -11, \dots)$, we have $R_1 = 1$, $R_2 = 2$, $R_3 = 2$, $R_4 = 2$, $R_5 = 1$, $R_6 = 5$, and $R_7 = 7$. Observe that rank order, from lowest to highest, of $(X_1, X_2, X_3, X_4, X_5)$ is $(X_2, X_3, X_4, X_1, X_5)$. This can be determined by the values of R_i for $i = 1, 2, \dots, 5$ as follows. Since $R_5 = 1$, we must have X_5 in the last slot, i.e. $(*, *, *, *, X_5)$. Since $R_4 = 2$, we know out of the remaining slots, X_4 must be in the second from the far most right, i.e. $(*, *, X_4, *, X_5)$. Since $R_3 = 2$, we know that X_3 is again the second from the right of the remaining slots, i.e. we now know, $(*, X_3, X_4, *, X_5)$. Similarly, $R_2 = 2$ implies $(X_2, X_3, X_4, *, X_5)$ and finally $R_1 = 1$ gives, $(X_2, X_3, X_4, X_1, X_5)$ $(= (-8, 4, 7, 9, 23)$ in the example). As another example, if $R_i = i$ for $i = 1, 2, \dots, n$, then $X_n < X_{n-1} < \cdots < X_1$.

Theorem 10.21 (Renyi Theorem). *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. and assume that $F(x) := P(X_n \le x)$ is continuous. The $\{R_n\}_{n=1}^\infty$ is an independent sequence,*

$$P(R_n = k) = \frac{1}{n} \text{ for } k = 1, 2, \dots, n,$$

and the events, $A_n = \{X_n \text{ is a record}\} = \{R_n = 1\}$ are independent as n varies and

$$P(A_n) = P(R_n = 1) = \frac{1}{n}.$$

Proof. By Problem 6 on p. 110 of Resnick or by Fubini's theorem, $(X_1, \dots, X_n)$ and $(X_{\sigma 1}, \dots, X_{\sigma n})$ have the same distribution for any permutation σ.

Since F is continuous, it now follows that up to a set of measure zero,

$$\Omega = \sum_\sigma \{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\}$$

and therefore

$$1 = P(\Omega) = \sum_\sigma P(\{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\}).$$

Since $P(\{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\})$ is independent of σ we may now conclude that

$$P(\{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\}) = \frac{1}{n!}$$

for all σ. As observed before the statement of the theorem, to each realization $(\varepsilon_1, \dots, \varepsilon_n)$, (here $\varepsilon_i \in \mathbb{N}$ with $\varepsilon_i \le i$) of $(R_1, \dots, R_n)$ there is a uniquely determined permutation, $\sigma = \sigma(\varepsilon_1, \dots, \varepsilon_n)$, such that $X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}$. (Notice that there are $n!$ permutations of $\{1, 2, \dots, n\}$ and there are also $n!$ choices for the $\{(\varepsilon_1, \dots, \varepsilon_n) : 1 \le \varepsilon_i \le i\}$.) From this it follows that

$$\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\} = \{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\}$$

and therefore,

$$P(\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\}) = P(X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}) = \frac{1}{n!}.$$

Since

$$\begin{aligned} P(\{R_n = \varepsilon_n\}) &= \sum_{(\varepsilon_1, \dots \varepsilon_{n-1})} P(\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\}) \\ &= \sum_{(\varepsilon_1, \dots \varepsilon_{n-1})} \frac{1}{n!} = (n-1)! \cdot \frac{1}{n!} = \frac{1}{n} \end{aligned}$$

we have shown that

$$P(\{(R_1, \dots, R_n) = (\varepsilon_1, \dots, \varepsilon_n)\}) = \frac{1}{n!} = \prod_{j=1}^n \frac{1}{j} = \prod_{j=1}^n P(\{R_j = \varepsilon_j\}).$$

■

10.3 Gaussian Random Vectors

As you saw in Exercise 10.2, uncorrelated random variables are typically not independent. However, if the random variables involved are jointly Gaussian, then independence and uncorrelated are actually the same thing!

Lemma 10.22. *Suppose that $Z = (X, Y)^{\mathrm{tr}}$ is a Gaussian random vector with $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^l$. Then X is independent of Y iff $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$. This lemma also holds more generally. Namely if $\{X^l\}_{l=1}^n$ is a sequence of random vectors such that $(X^1, \dots, X^n)$ is a Gaussian random vector. Then $\{X^l\}_{l=1}^n$ are independent iff $\operatorname{Cov}\left(X_i^l, X_k^{l'}\right) = 0$ for all $l \ne l'$ and i and k.*

Proof. We know by Exercise 10.2 that if X_i and Y_j are independent, then $\operatorname{Cov}(X_i, Y_j) = 0$. For the converse direction, if $\operatorname{Cov}(X_i, Y_j) = 0$ for all $1 \le i \le k$ and $1 \le j \le l$ and $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned} \operatorname{Var}(x \cdot X + y \cdot Y) &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y) + 2 \operatorname{Cov}(x \cdot X, y \cdot Y) \\ &= \operatorname{Var}(x \cdot X) + \operatorname{Var}(y \cdot Y). \end{aligned}$$

Therefore using the fact that (X, Y) is a Gaussian random vector,

$$\begin{aligned} \mathbb{E}\left[e^{ix \cdot X} e^{iy \cdot Y}\right] &= \mathbb{E}\left[e^{i(x \cdot X + y \cdot Y)}\right] \\ &= \exp\left(-\frac{1}{2} \operatorname{Var}(x \cdot X + y \cdot Y) + \mathbb{E}(x \cdot X + y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2} \operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X) - \frac{1}{2} \operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \mathbb{E}\left[e^{ix \cdot X}\right] \cdot \mathbb{E}\left[e^{iy \cdot Y}\right], \end{aligned}$$

and because x and y were arbitrary, we may conclude from Theorem 10.11 that X and Y are independent. ■

Corollary 10.23. *Suppose that $X : \Omega \to \mathbb{R}^k$ and $Y : \Omega \to \mathbb{R}^l$ are two independent random Gaussian vectors, then (X, Y) is also a Gaussian random vector. This corollary generalizes to multiple independent random Gaussian vectors.*

Proof. Let $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^l$, then

$$\begin{aligned} \mathbb{E}\left[e^{i(x,y) \cdot (X,Y)}\right] &= \mathbb{E}\left[e^{i(x \cdot X + y \cdot Y)}\right] = \mathbb{E}\left[e^{ix \cdot X} e^{iy \cdot Y}\right] = \mathbb{E}\left[e^{ix \cdot X}\right] \cdot \mathbb{E}\left[e^{iy \cdot Y}\right] \\ &= \exp\left(-\frac{1}{2} \operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X)\right) \\ &\quad \times \exp\left(-\frac{1}{2} \operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2} \operatorname{Var}(x \cdot X) + i\mathbb{E}(x \cdot X) - \frac{1}{2} \operatorname{Var}(y \cdot Y) + i\mathbb{E}(y \cdot Y)\right) \\ &= \exp\left(-\frac{1}{2} \operatorname{Var}(x \cdot X + y \cdot Y) + i\mathbb{E}(x \cdot X + y \cdot Y)\right) \end{aligned}$$

which shows that (X, Y) is again Gaussian. ■

Notation 10.24 *Suppose that* $\{X_i\}_{i=1}^n$ *is a collection of* $\mathbb{R}$ *– valued variables or* $\mathbb{R}^d$ *– valued random vectors. We will write* $X_1 \overset{\perp\perp}{+} X_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} X_n$ *for* $X_1 + \dots + X_n$ *under the additional assumption that the* $\{X_i\}_{i=1}^n$ *are independent.*

Corollary 10.25. *Suppose that* $\{X_i\}_{i=1}^n$ *are independent Gaussian random variables, then* $S_n := \sum_{i=1}^n X_i$ *is a Gaussian random variables with :*

$$\operatorname{Var}(S_n) = \sum_{i=1}^n \operatorname{Var}(X_i) \text{ and } \mathbb{E}S_n = \sum_{i=1}^n \mathbb{E}X_i, \tag{10.7}$$

i.e.

$$X_1 \overset{\perp\perp}{+} X_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} X_n \overset{d}{=} N\left(\sum_{i=1}^n \operatorname{Var}(X_i), \sum_{i=1}^n \mathbb{E}X_i\right).$$

In particular if $\{X_i\}_{i=1}^\infty$ *are i.i.d. Gaussian random variables with* $\mathbb{E}X_i = \mu$ *and* $\sigma^2 = \operatorname{Var}(X_i)$, *then*

$$\frac{S_n}{n} - \mu \overset{d}{=} N\left(0, \frac{\sigma^2}{n}\right) \text{ and} \tag{10.8}$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{d}{=} N(0,1). \tag{10.9}$$

Equation (10.9) is a very special case of the central limit theorem while Eq. (10.8) leads to a very special case of the strong law of large numbers, see Corollary 10.26.

Proof. The fact that S_n, $\frac{S_n}{n} - \mu$, and $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ are all Gaussian follows from Corollary 10.25 and Lemma 9.36 or by direct calculation. The formulas for the variances and means of these random variables are routine to compute. ∎

Recall the first Borel Cantelli-Lemma 7.14 states that if $\{A_n\}_{n=1}^\infty$ are measurable sets, then

$$\sum_{n=1}^\infty P(A_n) < \infty \implies P(\{A_n \text{ i.o.}\}) = 0. \tag{10.10}$$

Corollary 10.26. *Let* $\{X_i\}_{i=1}^\infty$ *be i.i.d. Gaussian random variables with* $\mathbb{E}X_i = \mu$ *and* $\sigma^2 = \operatorname{Var}(X_i)$. *Then* $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ *a.s. and moreover for every* $\alpha < \frac{1}{2}$, *there exists* $N_\alpha : \Omega \to \mathbb{N} \cup \{\infty\}$, *such that* $P(N_\alpha = \infty) = 0$ *and*

$$\left|\frac{S_n}{n} - \mu\right| \le n^{-\alpha} \text{ for } n \ge N_\alpha.$$

In particular, $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ *a.s.*

Proof. Let $Z \overset{d}{=} N(0,1)$ so that $\frac{\sigma}{\sqrt{n}} Z \overset{d}{=} N\left(0, \frac{\sigma^2}{n}\right)$. From the Eq. (10.8) and Eq. (7.42),

$$P\left(\left|\frac{S_n}{n} - \mu\right| \ge \varepsilon\right) = P\left(\left|\frac{\sigma}{\sqrt{n}} Z\right| \ge \varepsilon\right) = P\left(|Z| \ge \frac{\sqrt{n}\varepsilon}{\sigma}\right)$$
$$\le \exp\left(-\frac{1}{2}\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right)^2\right) = \exp\left(-\frac{\varepsilon^2}{2\sigma^2} n\right).$$

Taking $\varepsilon = n^{-\alpha}$ with $1 - 2\alpha > 0$, it follows that

$$\sum_{n=1}^\infty P\left(\left|\frac{S_n}{n} - \mu\right| \ge n^{-\alpha}\right) \le \sum_{n=1}^\infty \exp\left(-\frac{1}{2\sigma^2} n^{1-2\alpha}\right) < \infty$$

and so by the first Borel-Cantelli lemma,

$$P\left(\left\{\left|\frac{S_n}{n} - \mu\right| \ge n^{-\alpha} \text{ i.o.}\right\}\right) = 0.$$

Therefore, P – a.s., $\left|\frac{S_n}{n} - \mu\right| \le n^{-\alpha}$ a.a., and in particular $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s. ∎

10.4 Summing independent random variables

Exercise 10.3. Suppose that $X \overset{d}{=} N(0, a^2)$ and $Y \overset{d}{=} N(0, b^2)$ and X and Y are independent. Show by direct computation using the formulas for the distributions of X and Y that $X + Y = N(0, a^2 + b^2)$.

Solution to Exercise (10.3). If $f : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function, then

$$\mathbb{E}[f(X+Y)] = \frac{1}{Z}\int_{\mathbb{R}^2} f(x+y)\, e^{-\frac{1}{2a^2}x^2} e^{-\frac{1}{2b^2}y^2} dx dy,$$

where $Z = 2\pi ab$. Let us make the change of variables, $(x, z) = (x, x+y)$ and observe that $dxdy = dxdz$ (you check). Therefore we have,

$$\mathbb{E}[f(X+Y)] = \frac{1}{Z}\int_{\mathbb{R}^2} f(z)\, e^{-\frac{1}{2a^2}x^2} e^{-\frac{1}{2b^2}(z-x)^2} dx dz$$

which shows, $\operatorname{Law}(X+Y)(dz) = \rho(z)\, dz$ where

$$\rho(z) = \frac{1}{Z}\int_{\mathbb{R}} e^{-\frac{1}{2a^2}x^2} e^{-\frac{1}{2b^2}(z-x)^2} dx. \tag{10.11}$$

Working the exponent, for any $c \in \mathbb{R}$, we have

$$\begin{aligned}\frac{1}{a^2}x^2 + \frac{1}{b^2}(z-x)^2 &= \frac{1}{a^2}x^2 + \frac{1}{b^2}\left(x^2 - 2xz + z^2\right)\\ &= \left(\frac{1}{a^2}+\frac{1}{b^2}\right)x^2 - \frac{2}{b^2}xz + \frac{1}{b^2}z^2\\ &= \left(\frac{1}{a^2}+\frac{1}{b^2}\right)\left[(x-cz)^2 + 2cxz - c^2z^2\right] - \frac{2}{b^2}xz + \frac{1}{b^2}z^2.\end{aligned}$$

Let us now choose (to complete the squares) c such that where c must be chosen so that

$$c\left(\frac{1}{a^2}+\frac{1}{b^2}\right) = \frac{1}{b^2} \implies c = \frac{a^2}{a^2+b^2},$$

in which case,

$$\frac{1}{a^2}x^2 + \frac{1}{b^2}(z-x)^2 = \left(\frac{1}{a^2}+\frac{1}{b^2}\right)\left[(x-cz)^2\right] + \left[\frac{1}{b^2} - c^2\left(\frac{1}{a^2}+\frac{1}{b^2}\right)\right]z^2$$

where,

$$\frac{1}{b^2} - c^2\left(\frac{1}{a^2}+\frac{1}{b^2}\right) = \frac{1}{b^2}(1-c) = \frac{1}{a^2+b^2}.$$

So making the change of variables, $x \to x - cz$, in the integral in Eq. (10.11) implies,

$$\begin{aligned}\rho(z) &= \frac{1}{Z}\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\left(\frac{1}{a^2}+\frac{1}{b^2}\right)w^2 - \frac{1}{2}\frac{1}{a^2+b^2}z^2\right)dw\\ &= \frac{1}{\tilde{Z}}\exp\left(-\frac{1}{2}\frac{1}{a^2+b^2}z^2\right)\end{aligned}$$

where,

$$\begin{aligned}\frac{1}{\tilde{Z}} &= \frac{1}{Z}\cdot\int_{\mathbb{R}}\exp\left(-\frac{1}{2}\left(\frac{1}{a^2}+\frac{1}{b^2}\right)w^2\right)dw = \frac{1}{2\pi ab}\sqrt{2\pi\left(\frac{1}{a^2}+\frac{1}{b^2}\right)^{-1}}\\ &= \frac{1}{2\pi ab}\sqrt{2\pi\frac{a^2b^2}{a^2+b^2}} = \frac{1}{\sqrt{2\pi\left(a^2+b^2\right)}}.\end{aligned}$$

Thus it follows that $X \overset{\perp\perp}{+} Y \overset{d}{=} N\left(a^2+b^2, 0\right)$.

Exercise 10.4. Show that the sum, $N_1 + N_2$, of two independent Poisson random variables, N_1 and N_2, with parameters λ_1 and λ_2 respectively is again a Poisson random variable with parameter $\lambda_1 + \lambda_2$. (You could use generating functions or do this by hand.) In short $\mathrm{Poi}(\lambda_1) \overset{\perp\perp}{+} \mathrm{Poi}(\lambda_2) \overset{d}{=} \mathrm{Poi}(\lambda_1+\lambda_2)$.

Solution to Exercise (10.4). Let $z \in \mathbb{C}$, then by independence,

$$\begin{aligned}\mathbb{E}\left[z^{N_1+N_2}\right] &= \mathbb{E}\left[z^{N_1}z^{N_2}\right] = \mathbb{E}\left[z^{N_1}\right]\mathbb{E}\left[z^{N_2}\right]\\ &= e^{\lambda_1(z-1)}\cdot e^{\lambda_2(z-1)} = e^{(\lambda_1+\lambda_2)(z-1)}\end{aligned}$$

from which it follows that $N_1 + N_2 \overset{d}{=} \mathrm{Poisson}(\lambda_1+\lambda_2)$.

Example 10.27 (Gamma Distribution Sums). We will show here that $\mathrm{Gamma}(k,\theta) \overset{\perp\perp}{+} \mathrm{Gamma}(l,\theta) = \mathrm{Gamma}(k+l,\theta)$. In Exercise 7.13 you showed if $k, \theta > 0$ then

$$\mathbb{E}\left[e^{tX}\right] = (1-\theta t)^{-k} \text{ for } t < \theta^{-1}$$

where X is a positive random variable with $X \overset{d}{=} \mathrm{Gamma}(k,\theta)$, i.e.

$$(X_*P)(dx) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}dx \text{ for } x > 0.$$

Suppose that X and Y are independent Random variables with $X \overset{d}{=} \mathrm{Gamma}(k,\theta)$ and $Y \overset{d}{=} \mathrm{Gamma}(l,\theta)$ for some $l > 0$. It now follows that

$$\begin{aligned}\mathbb{E}\left[e^{t(X+Y)}\right] &= \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right]\\ &= (1-\theta t)^{-k}(1-\theta t)^{-l} = (1-\theta t)^{-(k+l)}.\end{aligned}$$

Therefore it follows from Exercise 8.2 that $X + Y \overset{d}{=} \mathrm{Gamma}(k+l,\theta)$.

Example 10.28 (Exponential Distribution Sums). If $\{T_k\}_{k=1}^n$ are independent random variables such that $T_k \overset{d}{=} E(\lambda_k)$ for all k, then

$$T_1 \overset{\perp\perp}{+} T_2 \overset{\perp\perp}{+} \dots \overset{\perp\perp}{+} T_n = \mathrm{Gamma}\left(n, \lambda^{-1}\right).$$

This follows directly from Example 10.27 using $E(\lambda) = \mathrm{Gamma}\left(1,\lambda^{-1}\right)$ and induction. We will verify this directly later on in Corollary **??**.

Example 10.27 may also be verified using brute force. To this end, suppose that $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function, then

$$\begin{aligned}\mathbb{E}\left[f(X+Y)\right] &= \int_{\mathbb{R}_+^2} f(x+y)\,x^{k-1}\frac{e^{-x/\theta}}{\theta^k\Gamma(k)}y^{l-1}\frac{e^{-y/\theta}}{\theta^l\Gamma(l)}dxdy\\ &= \frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)}\int_{\mathbb{R}_+^2} f(x+y)\,x^{k-1}y^{l-1}e^{-(x+y)/\theta}dxdy.\end{aligned}$$

Let us now make the change of variables, $x = x$ and $z = x + y$, so that $dxdy = dxdz$, to find,

$$\mathbb{E}[f(X+Y)] = \frac{1}{\theta^{k+l}\Gamma(k)\Gamma(l)} \int 1_{0\le x\le z<\infty} f(z)\, x^{k-1}(z-x)^{l-1} e^{-z/\theta} dxdz. \tag{10.12}$$

To finish the proof we must now do that x integral and show,

$$\int_0^z x^{k-1}(z-x)^{l-1} dx = z^{k+l-1}\frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}.$$

(In fact we already know this must be correct from our Laplace transform computations above.) First make the change of variable, $x = zt$ to find,

$$\int_0^z x^{k-1}(z-x)^{l-1} dx = z^{k+l-1} B(k,l)$$

where $B(k,l)$ is the **beta − function** defined by;

$$B(k,l) := \int_0^1 t^{k-1}(1-t)^{l-1} dt \text{ for } \operatorname{Re} k, \operatorname{Re} l > 0. \tag{10.13}$$

Combining these results with Eq. (10.12) then shows,

$$\mathbb{E}[f(X+Y)] = \frac{B(k,l)}{\theta^{k+l}\Gamma(k)\Gamma(l)} \int_0^\infty f(z)\, z^{k+l-1} e^{-z/\theta} dz. \tag{10.14}$$

Since we already know that

$$\int_0^\infty z^{k+l-1} e^{-z/\theta} dz = \theta^{k+l}\Gamma(k+l)$$

it follows by taking $f = 1$ in Eq. (10.14) that

$$1 = \frac{B(k,l)}{\theta^{k+l}\Gamma(k)\Gamma(l)}\theta^{k+l}\Gamma(k+l)$$

which implies,

$$B(k,l) = \frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}. \tag{10.15}$$

Therefore, using this back in Eq. (10.14) implies

$$\mathbb{E}[f(X+Y)] = \frac{1}{\theta^{k+l}\Gamma(k+l)} \int_0^\infty f(z)\, z^{k+l-1} e^{-z/\theta} dz$$

from which it follows that $X + Y \overset{d}{=} \text{Gamma}(k+l,\theta)$.

Let us pause to give a direct verification of Eq. (10.15). By definition of the gamma function,

$$\begin{aligned}\Gamma(k)\Gamma(l) &= \int_{\mathbb{R}_+^2} x^{k-1}e^{-x}y^{l-1}e^{-y}dxdy = \int_{\mathbb{R}_+^2} x^{k-1}y^{l-1}e^{-(x+y)}dxdy.\\ &= \int_{0\le x\le z<\infty} x^{k-1}(z-x)^{l-1}e^{-z}dxdz\end{aligned}$$

Making the change of variables, $x = x$ and $z = x + y$ it follows,

$$\Gamma(k)\Gamma(l) = \int_{0\le x\le z<\infty} x^{k-1}(z-x)^{l-1}e^{-z}dxdz.$$

Now make the change of variables, $x = zt$ to find,

$$\begin{aligned}\Gamma(k)\Gamma(l) &= \int_0^\infty dz e^{-z}\int_0^1 dt\,(zt)^{k-1}(z-tz)^{l-1} z\\ &= \int_0^\infty e^{-z}z^{k+l-1}dz \cdot \int_0^1 t^{k-1}(1-t)^{l-1}dt\\ &= \Gamma(k+l)\,B(k,l).\end{aligned}$$

Definition 10.29 (Beta distribution). *The β – distribution is*

$$d\mu_{x,y}(t) = \frac{t^{x-1}(1-t)^{y-1}dt}{B(x,y)}.$$

Observe that

$$\int_0^1 t\,d\mu_{x,y}(t) = \frac{B(x+1,y)}{B(x,y)} = \frac{\frac{\Gamma(x+1)\Gamma(y)}{\Gamma(x+y+1)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{x}{x+y}$$

and

$$\int_0^1 t^2 d\mu_{x,y}(t) = \frac{B(x+2,y)}{B(x,y)} = \frac{\frac{\Gamma(x+2)\Gamma(y)}{\Gamma(x+y+2)}}{\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}} = \frac{(x+1)\,x}{(x+y+1)(x+y)}.$$

10.5 A Strong Law of Large Numbers

Theorem 10.30 (A simple form of the strong law of large numbers). *If $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables such that $\mathbb{E}\left[|X_n|^4\right] < \infty$, then*

$$\lim_{n\to\infty}\frac{S_n}{n} = \mu \text{ a.s.}$$

where $S_n := \sum_{k=1}^n X_k$ and $\mu := \mathbb{E}X_n = \mathbb{E}X_1$.

Exercise 10.5. Use the following outline to give a proof of Theorem 10.30.

1. First show that $x^p \leq 1 + x^4$ for all $x \geq 0$ and $1 \leq p \leq 4$. Use this to conclude;
$$\mathbb{E}|X_n|^p \leq 1 + \mathbb{E}|X_n|^4 < \infty \text{ for } 1 \leq p \leq 4.$$
Thus $\gamma := \mathbb{E}\left[|X_n - \mu|^4\right]$ and the standard deviation (σ^2) of X_n defined by,
$$\sigma^2 := \mathbb{E}\left[X_n^2\right] - \mu^2 = \mathbb{E}\left[(X_n - \mu)^2\right] < \infty,$$
are finite constants independent of n.
2. Show for all $n \in \mathbb{N}$ that
$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^4\right] = \frac{1}{n^4}\left(n\gamma + 3n(n-1)\sigma^4\right) = \frac{1}{n^2}\left[n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4\right].$$
(Thus $\frac{S_n}{n} \to \mu$ in $L^4(P)$.)
3. Use item 2. and Chebyshev's inequality to show
$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4}{\varepsilon^4 n^2}.$$
4. Use item 3. and the first Borel Cantelli Lemma 7.14 to conclude $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s.

10.6 A Central Limit Theorem

In this section we will give a preliminary a couple versions of the central limit theorem following [7, Chapter 2.14]. Let us set up some notation. Given a square integrable random variable Y, let
$$\bar{Y} := \frac{Y - \mathbb{E}Y}{\sigma(Y)} \text{ where } \sigma(Y) := \sqrt{\mathbb{E}(Y - \mathbb{E}Y)^2} = \sqrt{\mathrm{Var}(Y)}.$$

Let us also recall that if $Z = N(0, \sigma^2)$, then $Z \overset{d}{=} \sqrt{\sigma}N(0,1)$ and so by Eq. (7.40) with $\beta = 3$ we have,
$$\mathbb{E}\left|Z^3\right| = \sigma^3\mathbb{E}|N(0,1)|^3 = \sqrt{8/\pi}\sigma^3. \tag{10.16}$$

Theorem 10.31 (A CLT proof w/o Fourier). *Suppose that* $\{X_k\}_{k=1}^\infty \subset L^3(P)$ *is a sequence of independent random variables such that*
$$C := \sup_k \mathbb{E}|X_k - \mathbb{E}X_k|^3 < \infty$$
Then for every function, $f \in C^3(\mathbb{R})$ *with* $M := \sup_{x\in\mathbb{R}}\left|f^{(3)}(x)\right| < \infty$ *we have*
$$\left|\mathbb{E}f(N) - \mathbb{E}f(\bar{S}_n)\right| \leq \frac{M}{3!}\left(1 + \sqrt{8/\pi}\right)\frac{C}{\sigma(S_n)^3}\cdot n, \tag{10.17}$$
where $S_n := X_1 + \dots + X_n$ *and* $N \overset{d}{=} N(0,1)$. *In particular if we further assume that*
$$\delta := \liminf_{n\to\infty}\frac{1}{n}\sigma(S_n)^2 = \liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{Var}(X_i) > 0, \tag{10.18}$$
Then it follows that
$$\left|\mathbb{E}f(N) - \mathbb{E}f(\bar{S}_n)\right| = O\left(\frac{1}{\sqrt{n}}\right) \text{ as } n \to \infty \tag{10.19}$$
which is to say, $\bar{S}_n$ *is "close" in distribution to* N, *which we abbreviate by* $\bar{S}_n \overset{d}{\cong} N$ *for large* n.

(It should be noted that the estimate in Eq. (10.17) is valid for any finite collection of random variables, $\{X_k\}_{k=1}^n$.*)*

Proof. Let $n \in \mathbb{N}$ be fixed and then Let $\{Y_k, N_k\}_{k=1}^\infty$ be a collection of independent random variables such that
$$Y_k \overset{d}{=} \bar{X}_k = \frac{X_k - \mathbb{E}X_k}{\sigma(S_n)} \text{ and } N_k \overset{d}{=} N(0, \mathrm{Var}(Y_k)) \text{ for } 1 \leq k \leq n.$$

Let $S_n^Y = Y_1 + \dots + Y_n \overset{d}{=} \bar{S}_n$ and $T_n := N_1 + \dots + N_n$. Since
$$\sum_{k=1}^n \mathrm{Var}(N_k) = \sum_{k=1}^n \mathrm{Var}(Y_k) = \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \mathrm{Var}(X_k - \mathbb{E}X_k) = \frac{1}{\sigma(S_n)^2}\sum_{k=1}^n \mathrm{Var}(X_k) = 1,$$

it follows by Corollary 10.25) that $T_n \overset{d}{=} N(0,1)$.

To compare $\mathbb{E}f(\bar{S}_n)$ with $\mathbb{E}f(N)$ we may compare $\mathbb{E}f(S_n^Y)$ with $\mathbb{E}f(T_n)$ which we will do by interpolating between S_n^Y and T_n. To this end, for $0 \leq k \leq n$, let

$$V_k := N_1 + \cdots + N_k + Y_{k+1} + \cdots + Y_n$$

with the convention that $V_n = T_n$ and $V_0 = S_n^Y$. Then by a telescoping series argument, it follows that

$$f(T_n) - f\left(S_n^Y\right) = f(V_n) - f(V_0) = \sum_{k=1}^{n} \left[f(V_k) - f(V_{k-1})\right]. \tag{10.20}$$

We now make use of Taylor's theorem with integral remainder the form,

$$f(x+\Delta) - f(x) = f'(x)\,\Delta + \frac{1}{2} f''(x)\,\Delta^2 + r(x,\Delta)\,\Delta^3 \tag{10.21}$$

where

$$r(x,\Delta) := \frac{1}{2}\int_0^1 f'''(x+t\Delta)(1-t)^2\,dt.$$

Taking Eq. (10.20) with Δ replaced by δ and subtracting the results then implies

$$f(x+\Delta) - f(x+\delta) = f'(x)(\Delta-\delta) + \frac{1}{2} f''(x)\left(\Delta^2 - \delta^2\right) + \rho(x,\Delta,\delta), \tag{10.22}$$

where

$$|\rho(x,\Delta,\delta)| = \left|r(x,\Delta)\,\Delta^3 - r(x,\delta)\,\delta^3\right| \le \frac{M}{3!}\left[|\Delta|^3 + |\delta|^3\right], \tag{10.23}$$

wherein we have used the simple estimate, $|r(x,\Delta)| \vee |r(x,\delta)| \le M/3!$.

If we define

$$U_k := N_1 + \cdots + N_{k-1} + Y_{k+1} + \cdots + Y_n,$$

then $V_k = U_k + N_k$ and $V_{k-1} = U_k + Y_k$. Hence, using Eq. (**??**) with $x = U_k$, $\Delta = N_k$and $\delta = Y_k$, it follows that

$$\begin{aligned} f(V_k) - f(V_{k-1}) &= f(U_k + N_k) - f(U_k + Y_k) \\ &= f'(U_k)(N_k - Y_k) + \frac{1}{2} f''(U_k)\left(N_k^2 - Y_k^2\right) + R_k \end{aligned} \tag{10.24}$$

where

$$|R_k| \le \frac{M}{3!}\left[|N_k|^3 + |Y_k|^3\right]. \tag{10.25}$$

Taking expectations of Eq. (10.24) using; Eq. (10.25), $\mathbb{E}N_k = 0 = \mathbb{E}Y_k$, $\mathbb{E}N_k^2 = \mathbb{E}Y_k^2$, and the fact that U_k is independent of both Y_k and N_k, we find

$$|\mathbb{E}[f(V_k) - f(V_{k-1})]| = |\mathbb{E}R_k| \le \frac{M}{3!}\mathbb{E}\left[|N_k|^3 + |Y_k|^3\right].$$

Making use of Eq. (10.16) it follows that

$$\mathbb{E}|N_k|^3 = \sqrt{8/\pi}\cdot\mathrm{Var}(N_k)^{3/2} = \sqrt{8/\pi}\cdot\mathrm{Var}(Y_k)^{3/2} = \sqrt{8/\pi}\cdot\left(\mathbb{E}Y_k^2\right)^{3/2} \le \sqrt{8/\pi}\cdot\mathbb{E}|Y_k|^3,$$

wherein we have used Jensen's (or Hölder's) inequality (see Chapter 12 below) for the last inequality. Combining these estimates with Eq. (10.20) shows,

$$\begin{aligned} \left|\mathbb{E}\left[f(T_n) - f\left(S_n^Y\right)\right]\right| &= \left|\sum_{k=1}^n \mathbb{E}R_k\right| \le \sum_{k=1}^n \mathbb{E}|R_k| \\ &\le \frac{M}{3!}\sum_{k=1}^n \mathbb{E}\left[|N_k|^3 + |Y_k|^3\right] \\ &\le \frac{M}{3!}\left(1+\sqrt{8/\pi}\right)\sum_{k=1}^n \mathbb{E}\left[|Y_k|^3\right]. \end{aligned} \tag{10.26}$$

Since

$$\mathbb{E}|Y_k|^3 = \mathbb{E}\left|\left(\frac{X_k - \mathbb{E}X_k}{\sigma(S_n)}\right)^3\right| \le \frac{C}{\sigma(S_n)^3} \text{ and}$$
$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| = \left|\mathbb{E}\left[f(T_n) - f\left(S_n^Y\right)\right]\right|,$$

we see that Eq. (10.17) now follows from Eq. (10.26). ■

Corollary 10.32. *Suppose that* $\{X_n\}_{n=1}^\infty$ *is a sequence of i.i.d. random variables in* $L^3(P)$, $C := \mathbb{E}|X_1 - \mathbb{E}X_1|^3 < \infty$, $S_n := X_1 + \cdots + X_n$, *and* $N \overset{d}{=} N(0,1)$. *Then for every function,* $f \in C^3(\mathbb{R})$ *with* $M := \sup_{x\in\mathbb{R}}\left|f^{(3)}(x)\right| < \infty$ *we have*

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \frac{M}{3!\sqrt{n}}\left(1+\sqrt{8/\pi}\right)\frac{C}{\mathrm{Var}(X_1)^{3/2}}. \tag{10.27}$$

(This is a specialized form of the "Berry–Esseen theorem.")

By a slight modification of the proof of Theorem 10.31 we have the following central limit theorem.

Theorem 10.33 (A CLT proof w/o Fourier). *Suppose that* $\{X_n\}_{n=1}^\infty$ *is a sequence of i.i.d. random variables in* $L^2(P)$, $S_n := X_1 + \cdots + X_n$, *and* $N \overset{d}{=} N(0,1)$. *Then for every function,* $f \in C^2(\mathbb{R})$ *with* $M := \sup_{x\in\mathbb{R}}\left|f^{(2)}(x)\right| < \infty$ *and* f'' *being uniformly continuous on* $\mathbb{R}$ *we have,*

$$\lim_{n\to\infty}\mathbb{E}f\left(\bar{S}_n\right) = \mathbb{E}f(N).$$

Proof. In this proof we use the following form of Taylor's theorem;

$$f(x+\Delta) - f(x) = f'(x)\,\Delta + \frac{1}{2} f''(x)\,\Delta^2 + r(x,\Delta)\,\Delta^2 \tag{10.28}$$

where

$$r(x,\Delta) = \int_0^1 \left[f''(x+t\Delta) - f''(x)\right](1-t)\,dt.$$

Taking Eq. (10.28) with Δ replaced by δ and subtracting the results then implies

$$f(x+\Delta) - f(x+\delta) = f'(x)(\Delta-\delta) + \frac{1}{2} f''(x)\left(\Delta^2 - \delta^2\right) + \rho(x,\Delta,\delta)$$

where now,

$$\rho(x,\Delta,\delta) = r(x,\Delta)\,\Delta^2 - r(x,\delta)\,\delta^2.$$

Since f'' is uniformly continuous it follows that

$$\varepsilon(\Delta) := \frac{1}{2}\sup\left\{|f''(x+t\Delta) - f(x)| : x \in \mathbb{R} \text{ and } 0 \le t \le 1\right\} \to 0$$

Thus we may conclude that

$$|r(x,\Delta)| \le \int_0^1 |f''(x+t\Delta) - f''(x)|(1-t)\,dt \le \int_0^1 2\varepsilon(\Delta)(1-t)\,dt = \varepsilon(\Delta).$$

and therefore that

$$|\rho(x,\Delta,\delta)| \le \varepsilon(\Delta)\,\Delta^2 + \varepsilon(\delta)\,\delta^2.$$

So working just as in the proof of Theorem 10.31 we may conclude,

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \sum_{k=1}^n \mathbb{E}|R_k|$$

where now,

$$|R_k| = \varepsilon(N_k)\,N_k^2 + \varepsilon(Y_k)\,Y_k^2.$$

Since the $\{Y_k\}_{k=1}^n$ and the $\{N_k\}_{k=1}^n$ are *i.i.d.* now it follows that

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le n \cdot \mathbb{E}\left[\varepsilon(N_1)\,N_1^2 + \varepsilon(Y_1)\,Y_1^2\right].$$

Since $\mathrm{Var}(S_n) = n \cdot \mathrm{Var}(X_1)$, we have $Y_1 = \frac{X_1 - \mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}$, $\mathrm{Var}(N_1) = \mathrm{Var}(Y_1) = \frac{1}{n}$ and therefore $N_1 \overset{d}{=} \sqrt{\frac{1}{n}}N$. Combining these observations shows,

$$\left|\mathbb{E}f(N) - \mathbb{E}f\left(\bar{S}_n\right)\right| \le \mathbb{E}\left[\varepsilon\left(\sqrt{\frac{1}{n}}N\right)N^2 + \varepsilon\left(\frac{X_1 - \mathbb{E}X_1}{\sqrt{n}\sigma(X_1)}\right)\frac{(X_1 - \mathbb{E}X_1)^2}{\sigma^2(X_1)}\right]$$

which goes to zero as $n \to \infty$ by the DCT. ■

Lemma 10.34. *Suppose that $\{W\} \cup \{W_n\}_{n=1}^\infty$ is a collection of random variables such that $\lim_{n\to\infty} \mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all $f \in C_c^\infty(\mathbb{R})$, then $\lim_{n\to\infty} \mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all bounded continuous functions, $f : \mathbb{R} \to \mathbb{R}$.*

Proof. According to Theorem **??** below it suffices to show $\lim_{n\to\infty} \mathbb{E}f(W_n) = \mathbb{E}f(W)$ for all $f \in C_c(\mathbb{R})$. For such a function, $f \in C_c(\mathbb{R})$, we may find[1] $f_k \in C_c^\infty(\mathbb{R})$ with all supports being contained in a compact subset of $\mathbb{R}$ such that $\varepsilon_k := \sup_{x\in\mathbb{R}} |f(x) - f_k(x)| \to 0$ as $k \to \infty$. We then have,

$$\begin{aligned} |\mathbb{E}f(W) - \mathbb{E}f(W_n)| &\le |\mathbb{E}f(W) - \mathbb{E}f_k(W)| + |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| + |\mathbb{E}f_k(W_n) - \mathbb{E}f \\ &\le \mathbb{E}|f(W) - f_k(W)| + |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| + \mathbb{E}|f_k(W_n) - f(W \\ &\le 2\varepsilon_k + |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)|. \end{aligned}$$

Therefore it follows that

$$\begin{aligned} \limsup_{n\to\infty} |\mathbb{E}f(W) - \mathbb{E}f(W_n)| &\le 2\varepsilon_k + \limsup_{n\to\infty} |\mathbb{E}f_k(W) - \mathbb{E}f_k(W_n)| \\ &= 2\varepsilon_k \to 0 \text{ as } k \to \infty. \end{aligned}$$

■

Corollary 10.35. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of independent random variables, then under the hypothesis on this sequence in either of Theorem 10.31 or Theorem 10.33 we have that $\lim_{n\to\infty} \mathbb{E}f\left(\bar{S}_n\right) = \mathbb{E}f(N(0,1))$ for all $f : \mathbb{R} \to \mathbb{R}$ which are bounded and continuous.*

For more on the methods employed in this section the reader is advised to look up "Stein's method." In Chapters **??** and **??** below, we will relax the assumptions in the above theorem. The proofs later will be based in the characteristic functional or equivalently the Fourier transform.

10.7 The Second Borel-Cantelli Lemma

Lemma 10.36. *If $0 \le x \le \frac{1}{2}$, then*

$$e^{-2x} \le 1 - x \le e^{-x}. \tag{10.29}$$

Moreover, the upper bound in Eq. (10.29) is valid for all $x \in \mathbb{R}$.

Proof. The upper bound follows by the convexity of e^{-x}, see Figure 10.1. For the lower bound we use the convexity of $\varphi(x) = e^{-2x}$ to conclude that the line joining $(0,1) = (0,\varphi(0))$ and $\left(1/2, e^{-1}\right) = (1/2, \varphi(1/2))$ lies above $\varphi(x)$ for $0 \le x \le 1/2$. Then we use the fact that the line $1 - x$ lies above this line

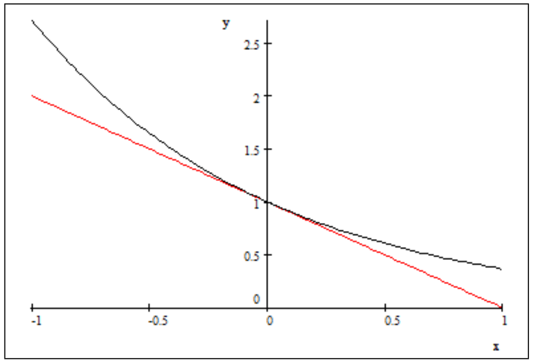


Fig. 10.1. A graph of $1-x$ and e^{-x} showing that $1-x \leq e^{-x}$ for all x.

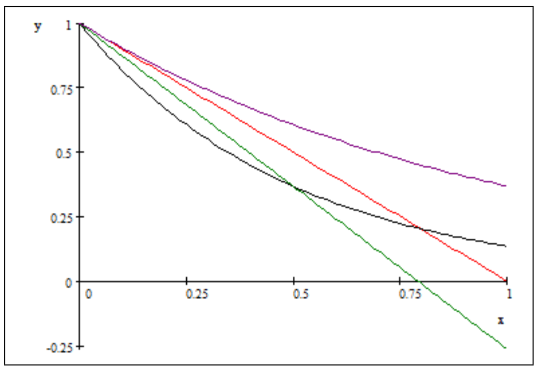


Fig. 10.2. A graph of $1-x$ (in red), the line joining $(0,1)$ and $\left(1/2, e^{-1}\right)$ (in green), e^{-x} (in purple), and e^{-2x} (in black) showing that $e^{-2x} \leq 1-x \leq e^{-x}$ for all $x \in [0,1/2]$.

to conclude the lower bound in Eq. (10.29), see Figure 10.2. See Example 12.46 below for a more formal proof of this lemma. ■

For $\{a_n\}_{n=1}^\infty \subset [0,1]$, let

$$\prod_{n=1}^{\infty}(1-a_n) := \lim_{N\to\infty}\prod_{n=1}^{N}(1-a_n).$$

The limit exists since, $\prod_{n=1}^{N}(1-a_n)$ decreases as N increases.

Exercise 10.6. Show; if $\{a_n\}_{n=1}^\infty \subset [0,1)$, then

$$\prod_{n=1}^{\infty}(1-a_n) = 0 \iff \sum_{n=1}^{\infty} a_n = \infty.$$

The implication, $\Longleftarrow$, holds even if $a_n = 1$ is allowed.

Solution to Exercise (10.6). By Eq. (10.29) we always have,

$$\prod_{n=1}^{N}(1-a_n) \leq \prod_{n=1}^{N} e^{-a_n} = \exp\left(-\sum_{n=1}^{N} a_n\right)$$

which upon passing to the limit as $N \to \infty$ gives

$$\prod_{n=1}^{\infty}(1-a_n) \leq \exp\left(-\sum_{n=1}^{\infty} a_n\right).$$

Hence if $\sum_{n=1}^\infty a_n = \infty$ then $\prod_{n=1}^\infty (1-a_n) = 0$.

Conversely, suppose that $\sum_{n=1}^\infty a_n < \infty$. In this case $a_n \to 0$ as $n \to \infty$ and so there exists an $m \in \mathbb{N}$ such that $a_n \in [0,1/2]$ for all $n \geq m$. Therefore by Eq. (10.29), for any $N \geq m$,

$$\begin{aligned}\prod_{n=1}^{N}(1-a_n) &= \prod_{n=1}^{m}(1-a_n)\cdot\prod_{n=m+1}^{N}(1-a_n)\\ &\geq \prod_{n=1}^{m}(1-a_n)\cdot\prod_{n=m+1}^{N}e^{-2a_n} = \prod_{n=1}^{m}(1-a_n)\cdot\exp\left(-2\sum_{n=m+1}^{N}a_n\right)\\ &\geq \prod_{n=1}^{m}(1-a_n)\cdot\exp\left(-2\sum_{n=m+1}^{\infty}a_n\right).\end{aligned}$$

So again letting $N \to \infty$ shows,

$$\prod_{n=1}^{\infty}(1-a_n) \geq \prod_{n=1}^{m}(1-a_n)\cdot\exp\left(-2\sum_{n=m+1}^{\infty}a_n\right) > 0.$$

Lemma 10.37 (Second Borel-Cantelli Lemma). *Suppose that* $\{A_n\}_{n=1}^\infty$ *are independent sets. If*

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \tag{10.30}$$

then

$$P(\{A_n \ i.o.\}) = 1. \tag{10.31}$$

Combining this with the first Borel Cantelli Lemma 7.14 gives the ***(Borel) Zero-One law****,*

[1] We will eventually prove this standard real analysis fact later in the course.

$$P(A_n \ i.o.) = \begin{cases} 0 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 & \text{if} \quad \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}.$$

Proof. We are going to prove Eq. (10.31) by showing,

$$0 = P(\{A_n \text{ i.o.}\}^c) = P(\{A_n^c \text{ a.a}\}) = P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c).$$

Since $\cap_{k \geq n} A_k^c \uparrow \cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c$ as $n \to \infty$ and $\cap_{k=n}^{m} A_k^c \downarrow \cap_{n=1}^{\infty} \cup_{k \geq n} A_k$ as $m \to \infty$,

$$P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c) = \lim_{n \to \infty} P(\cap_{k \geq n} A_k^c) = \lim_{n \to \infty} \lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c).$$

Making use of the independence of $\{A_k\}_{k=1}^{\infty}$ and hence the independence of $\{A_k^c\}_{k=1}^{\infty}$, we have

$$P(\cap_{m \geq k \geq n} A_k^c) = \prod_{m \geq k \geq n} P(A_k^c) = \prod_{m \geq k \geq n} (1 - P(A_k)). \tag{10.32}$$

Using the upper estimate in Eq. (10.29) along with Eq. (10.32) shows

$$P(\cap_{m \geq k \geq n} A_k^c) \leq \prod_{m \geq k \geq n} e^{-P(A_k)} = \exp\left(-\sum_{k=n}^{m} P(A_k)\right).$$

Using Eq. (10.30), we find from the above inequality that $\lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = 0$ and hence

$$P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c) = \lim_{n \to \infty} \lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = \lim_{n \to \infty} 0 = 0$$

Note: we could also appeal to Exercise 10.6 above to give a proof of the Borel Zero-One law without appealing to the first Borel Cantelli Lemma. ■

Example 10.38 (Example 7.15 continued). Suppose that $\{X_n\}$ are now independent Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. Then $P(\lim_{n \to \infty} X_n = 0) = 1$ iff $\sum p_n < \infty$. Indeed, $P(\lim_{n \to \infty} X_n = 0) = 1$ iff $P(X_n = 0 \text{ a.a.}) = 1$ iff $P(X_n = 1 \text{ i.o.}) = 0$ iff $\sum p_n = \sum P(X_n = 1) < \infty$.

Proposition 10.39 (Extremal behaviour of iid random variables). *Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables and c_n is an increasing sequence of positive real numbers such that for all $\alpha > 1$ we have*

$$\sum_{n=1}^{\infty} P\left(X_1 > \alpha^{-1} c_n\right) = \infty \tag{10.33}$$

while

$$\sum_{n=1}^{\infty} P(X_1 > \alpha c_n) < \infty. \tag{10.34}$$

Then

$$\limsup_{n \to \infty} \frac{X_n}{c_n} = 1 \ a.s. \tag{10.35}$$

Proof. By the second Borel-Cantelli Lemma, Eq. (10.33) implies

$$P\left(X_n > \alpha^{-1} c_n \text{ i.o. } n\right) = 1$$

from which it follows that

$$\limsup_{n \to \infty} \frac{X_n}{c_n} \geq \alpha^{-1} \text{ a.s..}$$

Taking $\alpha = \alpha_k = 1 + 1/k$, we find

$$P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} \geq 1\right) = P\left(\cap_{k=1}^{\infty} \left\{\limsup_{n \to \infty} \frac{X_n}{c_n} \geq \frac{1}{\alpha_k}\right\}\right) = 1.$$

Similarly, by the first Borel-Cantelli lemma, Eq. (10.34) implies

$$P(X_n > \alpha c_n \text{ i.o. } n) = 0$$

or equivalently,

$$P(X_n \leq \alpha c_n \text{ a.a. } n) = 1.$$

That is to say,

$$\limsup_{n \to \infty} \frac{X_n}{c_n} \leq \alpha \text{ a.s.}$$

and hence working as above,

$$P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} \leq 1\right) = P\left(\cap_{k=1}^{\infty} \left\{\limsup_{n \to \infty} \frac{X_n}{c_n} \leq \alpha_k\right\}\right) = 1.$$

Hence,

$$P\left(\limsup_{n \to \infty} \frac{X_n}{c_n} = 1\right) = P\left(\left\{\limsup_{n \to \infty} \frac{X_n}{c_n} \geq 1\right\} \cap \left\{\limsup_{n \to \infty} \frac{X_n}{c_n} \leq 1\right\}\right) = 1.$$

■

Example 10.40. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. standard normal random variables. Then by Mills' ratio (see Lemma 7.59),

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha c_n} e^{-\alpha^2 c_n^2/2}.$$

Now, suppose that we take c_n so that

$$e^{-c_n^2/2} = \frac{1}{n} \implies c_n = \sqrt{2\ln(n)}.$$

It then follows that

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\sqrt{2\pi}\alpha\sqrt{2\ln(n)}} e^{-\alpha^2 \ln(n)} = \frac{1}{2\alpha\sqrt{\pi\ln(n)}} \frac{1}{n^{-\alpha^2}}$$

and therefore

$$\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) = \infty \text{ if } \alpha < 1$$

and

$$\sum_{n=1}^{\infty} P(X_n \geq \alpha c_n) < \infty \text{ if } \alpha > 1.$$

Hence an application of Proposition 10.39 shows

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \text{ a.s..}$$

Example 10.41. Let $\{E_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with exponential distributions determined by

$$P(E_n > x) = e^{-(x\vee 0)} \text{ or } P(E_n \leq x) = 1 - e^{-(x\vee 0)}.$$

(Observe that $P(E_n \leq 0) = 0$) so that $E_n > 0$ a.s.) Then for $c_n > 0$ and $\alpha > 0$, we have

$$\sum_{n=1}^{\infty} P(E_n > \alpha c_n) = \sum_{n=1}^{\infty} e^{-\alpha c_n} = \sum_{n=1}^{\infty} \left(e^{-c_n}\right)^{\alpha}.$$

Hence if we choose $c_n = \ln n$ so that $e^{-c_n} = 1/n$, then we have

$$\sum_{n=1}^{\infty} P(E_n > \alpha \ln n) = \sum_{n=1}^{\infty} \left(\frac{1}{n}\right)^{\alpha}$$

which is convergent iff $\alpha > 1$. So by Proposition 10.39, it follows that

$$\limsup_{n\to\infty} \frac{E_n}{\ln n} = 1 \text{ a.s.}$$

Example 10.42. * Suppose now that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. distributed by the Poisson distribution with intensity, λ, i.e.

$$P(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In this case we have

$$P(X_1 \geq n) = e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} \geq \frac{\lambda^n}{n!} e^{-\lambda}$$

and

$$\begin{aligned}\sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=n}^{\infty} \frac{n!}{k!} \lambda^{k-n} \\ &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{n!}{(k+n)!} \lambda^k \leq \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k = \frac{\lambda^n}{n!}.\end{aligned}$$

Thus we have shown that

$$\frac{\lambda^n}{n!} e^{-\lambda} \leq P(X_1 \geq n) \leq \frac{\lambda^n}{n!}.$$

Thus in terms of convergence issues, we may assume that

$$P(X_1 \geq x) \sim \frac{\lambda^x}{x!} \sim \frac{\lambda^x}{\sqrt{2\pi x} e^{-x} x^x}$$

wherein we have used Stirling's formula,

$$x! \sim \sqrt{2\pi x} e^{-x} x^x.$$

Now suppose that we wish to choose c_n so that

$$P(X_1 \geq c_n) \sim 1/n.$$

This suggests that we need to solve the equation, $x^x = n$. Taking logarithms of this equation implies that

$$x = \frac{\ln n}{\ln x}$$

and upon iteration we find,

$$\begin{aligned}x &= \frac{\ln n}{\ln\left(\frac{\ln n}{\ln x}\right)} = \frac{\ln n}{\ell_2(n) - \ell_2(x)} = \frac{\ln n}{\ell_2(n) - \ell_2\left(\frac{\ln n}{\ln x}\right)} \\ &= \frac{\ln n}{\ell_2(n) - \ell_3(n) + \ell_3(x)}.\end{aligned}$$

where $\ell_k = \overbrace{\ln \circ \ln \circ \cdots \circ \ln}^{k\text{ - times}}$. Since, $x \leq \ln(n)$, it follows that $\ell_3(x) \leq \ell_3(n)$ and hence

$$x = \frac{\ln(n)}{\ell_2(n) + O(\ell_3(n))} = \frac{\ln(n)}{\ell_2(n)}\left(1 + O\left(\frac{\ell_3(n)}{\ell_2(n)}\right)\right).$$

Thus we are lead to take $c_n := \frac{\ln(n)}{\ell_2(n)}$. We then have, for $\alpha \in (0,\infty)$ that

$$\begin{aligned}(\alpha c_n)^{\alpha c_n} &= \exp(\alpha c_n[\ln\alpha + \ln c_n]) \\ &= \exp\left(\alpha\frac{\ln(n)}{\ell_2(n)}[\ln\alpha + \ell_2(n) - \ell_3(n)]\right) \\ &= \exp\left(\alpha\left[\frac{\ln\alpha - \ell_3(n)}{\ell_2(n)} + 1\right]\ln(n)\right) \\ &= n^{\alpha(1+\varepsilon_n(\alpha))}\end{aligned}$$

where

$$\varepsilon_n(\alpha) := \frac{\ln\alpha - \ell_3(n)}{\ell_2(n)}.$$

Hence we have

$$P(X_1 \ge \alpha c_n) \sim \frac{\lambda^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}e^{-\alpha c_n}(\alpha c_n)^{\alpha c_n}} \sim \frac{(\lambda/e)^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}}\frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}}.$$

Since

$$\ln(\lambda/e)^{\alpha c_n} = \alpha c_n \ln(\lambda/e) = \alpha\frac{\ln n}{\ell_2(n)}\ln(\lambda/e) = \ln n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}},$$

it follows that

$$(\lambda/e)^{\alpha c_n} = n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}}.$$

Therefore,

$$P(X_1 \ge \alpha c_n) \sim \frac{n^{\alpha\frac{\ln(\lambda/e)}{\ell_2(n)}}}{\sqrt{\frac{\ln(n)}{\ell_2(n)}}}\frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}} = \sqrt{\frac{\ell_2(n)}{\ln(n)}}\frac{1}{n^{\alpha(1+\delta_n(\alpha))}}$$

where $\delta_n(\alpha) \to 0$ as $n \to \infty$. From this observation, we may show,

$$\sum_{n=1}^{\infty} P(X_1 \ge \alpha c_n) < \infty \text{ if } \alpha > 1 \text{ and}$$
$$\sum_{n=1}^{\infty} P(X_1 \ge \alpha c_n) = \infty \text{ if } \alpha < 1$$

and so by Proposition 10.39 we may conclude that

$$\limsup_{n\to\infty}\frac{X_n}{\ln(n)/\ell_2(n)} = 1 \text{ a.s.}$$

10.8 Kolmogorov and Hewitt-Savage Zero-One Laws

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables on a measurable space, $(\Omega,\mathcal{B})$. Let $\mathcal{B}_n := \sigma(X_1,\dots,X_n)$, $\mathcal{B}_\infty := \sigma(X_1,X_2,\dots)$, $\mathcal{T}_n := \sigma(X_{n+1},X_{n+2},\dots)$, and $\mathcal{T} := \cap_{n=1}^{\infty}\mathcal{T}_n \subset \mathcal{B}_\infty$. We call $\mathcal{T}$ the **tail σ − field** and events, $A \in \mathcal{T}$, are called **tail events.**

Example 10.43. Let $S_n := X_1+\cdots+X_n$ and $\{b_n\}_{n=1}^{\infty} \subset (0,\infty)$ such that $b_n \uparrow \infty$. Here are some example of tail events and tail measurable random variables:

1. $\{\sum_{n=1}^{\infty} X_n \text{ converges}\} \in \mathcal{T}$. Indeed,
$$\left\{\sum_{k=1}^{\infty} X_k \text{ converges}\right\} = \left\{\sum_{k=n+1}^{\infty} X_k \text{ converges}\right\} \in \mathcal{T}_n$$
for all $n \in \mathbb{N}$.
2. Both $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are $\mathcal{T}$ − measurable as are $\limsup_{n\to\infty}\frac{S_n}{b_n}$ and $\liminf_{n\to\infty}\frac{S_n}{b_n}$.
3. $\{\lim X_n \text{ exists in } \bar{\mathbb{R}}\} = \left\{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\right\} \in \mathcal{T}$ and similarly,
$$\left\{\lim\frac{S_n}{b_n} \text{ exists in } \bar{\mathbb{R}}\right\} = \left\{\limsup_{n\to\infty}\frac{S_n}{b_n} = \liminf_{n\to\infty}\frac{S_n}{b_n}\right\} \in \mathcal{T}$$
and
$$\left\{\lim\frac{S_n}{b_n} \text{ exists in } \mathbb{R}\right\} = \left\{-\infty < \limsup_{n\to\infty}\frac{S_n}{b_n} = \liminf_{n\to\infty}\frac{S_n}{b_n} < \infty\right\} \in \mathcal{T}.$$
4. $\left\{\lim_{n\to\infty}\frac{S_n}{b_n} = 0\right\} \in \mathcal{T}$. Indeed, for any $k \in \mathbb{N}$,
$$\lim_{n\to\infty}\frac{S_n}{b_n} = \lim_{n\to\infty}\frac{(X_{k+1}+\cdots+X_n)}{b_n}$$
from which it follows that $\left\{\lim_{n\to\infty}\frac{S_n}{b_n} = 0\right\} \in \mathcal{T}_k$ for all k.

Definition 10.44. *Let $(\Omega,\mathcal{B},P)$ be a probability space. A σ – field, $\mathcal{F} \subset \mathcal{B}$ is **almost trivial** iff $P(\mathcal{F}) = \{0,1\}$, i.e. $P(A) \in \{0,1\}$ for all $A \in \mathcal{F}$.*

The following conditions on a sub-σ-algebra, $\mathcal{F} \subset \mathcal{B}$ are equivalent; 1) $\mathcal{F}$ is almost trivial, 2) $P(A) = P(A)^2$ for all $A \in \mathcal{F}$, and 3) $\mathcal{F}$ is independent of itself. For example if $\mathcal{F}$ is independent of itself, then $P(A) = P(A \cap A) = P(A)P(A)$ for all $A \in \mathcal{F}$ which implies $P(A) = 0$ or 1. If $\mathcal{F}$ is almost trivial and $A, B \in \mathcal{F}$, then $P(A \cap B) = 1 = P(A)P(B)$ if $P(A) = P(B) = 1$ and $P(A \cap B) = 0 = P(A)P(B)$ if either $P(A) = 0$ or $P(B) = 0$. Therefore $\mathcal{F}$ is independent of itself.

Lemma 10.45. *Suppose that $X : \Omega \to \bar{\mathbb{R}}$ is a random variable which is $\mathcal{F}$ measurable, where $\mathcal{F} \subset \mathcal{B}$ is almost trivial. Then there exists $c \in \bar{\mathbb{R}}$ such that $X = c$ a.s.*

Proof. Since $\{X = \infty\}$ and $\{X = -\infty\}$ are in $\mathcal{F}$, if $P(X = \infty) > 0$ or $P(X = -\infty) > 0$, then $P(X = \infty) = 1$ or $P(X = -\infty) = 1$ respectively. Hence, it suffices to finish the proof under the added condition that $P(X \in \mathbb{R}) = 1$.

For each $x \in \mathbb{R}$, $\{X \le x\} \in \mathcal{F}$ and therefore, $P(X \le x)$ is either 0 or 1. Since the function, $F(x) := P(X \le x) \in \{0,1\}$ is right continuous, non-decreasing and $F(-\infty) = 0$ and $F(+\infty) = 1$, there is a unique point $c \in \mathbb{R}$ where $F(c) = 1$ and $F(c-) = 0$. At this point, we have $P(X = c) = 1$.

Alternatively if $X : \Omega \to \mathbb{R}$ is an integrable $\mathcal{F}$ measurable random variable, we know that X is independent of itself and therefore X^2 is integrable and $\mathbb{E}X^2 = (\mathbb{E}X)^2 =: c^2$. Thus it follows that $\mathbb{E}\left[(X-c)^2\right] = 0$, i.e. $X = c$ a.s. For general $X : \Omega \to \mathbb{R}$, let $X_M := (M \wedge X) \vee (-M)$, then $X_M = \mathbb{E}X_M$ a.s. For sufficiently large M we know by MCT that $P(|X| < M) > 0$ and since $X = X_M = \mathbb{E}X_M$ a.s. on $\{|X| < M\}$, it follows that $c = \mathbb{E}X_M$ is constant independent of M for M large. Therefore, $X = \lim_{M\to\infty} X_M \overset{\text{a.s.}}{=} \lim_{M\to\infty} c = c$. ■

Proposition 10.46 (Kolmogorov's Zero-One Law). *Suppose that P is a probability measure on $(\Omega, \mathcal{B})$ such that $\{X_n\}_{n=1}^\infty$ are independent random variables. Then $\mathcal{T}$ is almost trivial, i.e. $P(A) \in \{0,1\}$ for all $A \in \mathcal{T}$. In particular the tail events in Example 10.43 have probability either 0 or 1.*

Proof. For each $n \in \mathbb{N}$, $\mathcal{T} \subset \sigma(X_{n+1}, X_{n+2}, \dots)$ which is independent of $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$. Therefore $\mathcal{T}$ is independent of $\cup\mathcal{B}_n$ which is a multiplicative system. Therefore $\mathcal{T}$ and is independent of $\mathcal{B}_\infty = \sigma(\cup\mathcal{B}_n) = \vee_{n=1}^\infty \mathcal{B}_n$. As $\mathcal{T} \subset \mathcal{B}_\infty$ it follows that $\mathcal{T}$ is independent of itself, i.e. $\mathcal{T}$ is almost trivial. ■

Corollary 10.47. *Keeping the assumptions in Proposition 10.46 and let $\{b_n\}_{n=1}^\infty \subset (0,\infty)$ such that $b_n \uparrow \infty$. Then $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$, $\limsup_{n\to\infty} \frac{S_n}{b_n}$, and $\liminf_{n\to\infty} \frac{S_n}{b_n}$ are all constant almost surely. In particular, either $P\left(\left\{\lim_{n\to\infty} \frac{S_n}{b_n} \text{ exists}\right\}\right) = 0$ or $P\left(\left\{\lim_{n\to\infty} \frac{S_n}{b_n} \text{ exists}\right\}\right) = 1$ and in the latter case $\lim_{n\to\infty} \frac{S_n}{b_n} = c$ a.s for some $c \in \bar{\mathbb{R}}$.*

Example 10.48. Suppose that $\{A_n\}_{n=1}^\infty$ are independent sets and let $X_n := 1_{A_n}$ for all n and $\mathcal{T} = \cap_{n\ge1}\sigma(X_n, X_{n+1}, \dots)$. Then $\{A_n \text{ i.o.}\} \in \mathcal{T}$ and therefore by the Kolmogorov 0-1 law, $P(\{A_n \text{ i.o.}\}) = 0$ or 1. Of course, in this case the Borel zero - one laws tells when $P(\{A_n \text{ i.o.}\})$ is 0 and when it is 1 depending on whether $\sum_{n=1}^\infty P(A_n)$ is finite or infinite respectively.

10.8.1 Hewitt-Savage Zero-One Law

In this subsection, let $\Omega := \mathbb{R}^\infty = \mathbb{R}^{\mathbb{N}}$ and $X_n(\omega) = \omega_n$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, and $\mathcal{B} := \sigma(X_1, X_2, \dots)$ be the product σ – algebra on Ω. We say a permutation (i.e. a bijective map on $\mathbb{N}$), $\pi : \mathbb{N} \to \mathbb{N}$ is finite if $\pi(n) = n$ for a.a. n. Define $T_\pi : \Omega \to \Omega$ by $T_\pi(\omega) = (\omega_{\pi1}, \omega_{\pi2}, \dots)$. Since $X_i \circ T_\pi(\omega) = \omega_{\pi i} = X_{\pi i}(\omega)$ for all i, it follows that T_π is $\mathcal{B}/\mathcal{B}$ – measurable.

Let us further suppose that μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and let $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $\left(\Omega = \mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$. Then $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables with $\text{Law}_P(X_n) = \mu$ for all n. If $\pi : \mathbb{N} \to \mathbb{N}$ is a finite permutation and $A_i \in \mathcal{B}_{\mathbb{R}}$ for all i, then

$$T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \dots) = A_{\pi^{-1}1} \times A_{\pi^{-1}2} \times \dots.$$

Since sets of the form, $A_1 \times A_2 \times A_3 \times \dots$, form a π – system generating $\mathcal{B}$ and

$$\begin{aligned} P \circ T_\pi^{-1}(A_1 \times A_2 \times A_3 \times \dots) &= \prod_{i=1}^\infty \mu(A_{\pi^{-1}i}) \\ &= \prod_{i=1}^\infty \mu(A_i) = P(A_1 \times A_2 \times A_3 \times \dots), \end{aligned}$$

we may conclude that $P \circ T_\pi^{-1} = P$.

Definition 10.49. *The **permutation invariant** σ – field, $\mathcal{S} \subset \mathcal{B}$, is the collection of sets, $A \in \mathcal{B}$ such that $T_\pi^{-1}(A) = A$ for all finite permutations π. (You should check that $\mathcal{S}$ is a σ – field!)*

Proposition 10.50 (Hewitt-Savage Zero-One Law). *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $P = \otimes_{n=1}^\infty \mu$ be the infinite product measure on $\left(\Omega = \mathbb{R}^{\mathbb{N}}, \mathcal{B}\right)$ so that $\{X_n\}_{n=1}^\infty$ (recall that $X_n(\omega) = \omega_n$) is an i.i.d. sequence with $\text{Law}_P(X_n) = \mu$ for all n. Then $\mathcal{S}$ is P – almost trivial.*

Proof. Let $B \in \mathcal{S}$, $f = 1_B$, and $g = G(X_1, \dots, X_n)$ be a $\sigma(X_1, X_2, \dots, X_n)$ – measurable function such that $\sup_{\omega\in\Omega}|g(\omega)| \le 1$. Further let π be a finite permutation such that $\{\pi1, \dots, \pi n\} \cap \{1, 2, \dots, n\} = \emptyset$ – for example we could take $\pi(j) = j + n$, $\pi(j+n) = j$ for $j = 1, 2, \dots, n$, and $\pi(j + 2n) = j + 2n$ for all $j \in \mathbb{N}$. Then $g \circ T_\pi = G(X_{\pi1}, \dots, X_{\pi n})$ is independent of g and therefore,

$$(\mathbb{E}g)^2 = \mathbb{E}g \cdot \mathbb{E}[g \circ T_\pi] = \mathbb{E}[g \cdot g \circ T_\pi].$$

Since $f \circ T_\pi = 1_{T_\pi^{-1}(B)} = 1_B = f$, it follows that $\mathbb{E}f = \mathbb{E}f^2 = \mathbb{E}[f \cdot f \circ T_\pi]$ and therefore,

$$\left|\mathbb{E}f - (\mathbb{E}g)^2\right| = |\mathbb{E}[f \cdot f \circ T_\pi - g \cdot g \circ T_\pi]|$$
$$\leq \mathbb{E}|[f-g]\, f \circ T_\pi| + \mathbb{E}|g[f\circ T_\pi - g \circ T_\pi]|$$
$$\leq \mathbb{E}|f-g| + \mathbb{E}|f\circ T_\pi - g\circ T_\pi| = 2\mathbb{E}|f-g|. \tag{10.36}$$

According to Corollary 8.13 (or see Corollary 5.28 or Theorem 5.44 or Exercise 8.5)), we may choose $g = g_k$ as above with $\mathbb{E}|f-g_k| \to 0$ as $n \to \infty$ and so passing to the limit in Eq. (10.36) with $g = g_k$, we may conclude,

$$\left|P(B) - [P(B)]^2\right| = \left|\mathbb{E}f - (\mathbb{E}f)^2\right| \leq 0.$$

That is $P(B) \in \{0,1\}$ for all $B \in \mathcal{S}$. ■

In a nutshell, here is the crux of the above proof. First off we know that for $B \in \mathcal{S} \subset \mathcal{B}$, there exists g which is $\sigma(X_1,\dots,X_n)$ – measurable such that $f := 1_B \cong g$. Since $P \circ T_\pi^{-1} = P$ it also follows that $f = f\circ T_\pi \cong g \circ T_\pi$. For judiciously chosen π, we know that g and $g \circ T_\pi$ are independent. Therefore

$$\mathbb{E}f^2 = \mathbb{E}[f\cdot f\circ T_\pi] \cong \mathbb{E}[g\cdot g\circ T_\pi] = \mathbb{E}[g]\cdot\mathbb{E}[g\circ T_\pi] = (\mathbb{E}g)^2 \cong (\mathbb{E}f)^2.$$

As the approximation f by g may be made as accurate as we please, it follows that $P(B) = \mathbb{E}f^2 = (\mathbb{E}f)^2 = [P(B)]^2$ for all $B \in \mathcal{S}$.

Example 10.51 (Some Random Walk $0-1$ Law Results). Continue the notation in Proposition 10.50.

1. As above, if $S_n = X_1 + \dots + X_n$, then $P(S_n \in B \text{ i.o.}) \in \{0,1\}$ for all $B \in \mathcal{B}_{\mathbb{R}}$. Indeed, if π is a finite permutation,
$$T_\pi^{-1}(\{S_n \in B \text{ i.o.}\}) = \{S_n \circ T_\pi \in B \text{ i.o.}\} = \{S_n \in B \text{ i.o.}\}.$$
Hence $\{S_n \in B \text{ i.o.}\}$ is in the permutation invariant σ – field, $\mathcal{S}$. The same goes for $\{S_n \in B \text{ a.a.}\}$
2. If $P(X_1 \neq 0) > 0$, then $\limsup_{n\to\infty} S_n = \infty$ a.s. or $\limsup_{n\to\infty} S_n = -\infty$ a.s. Indeed,
$$T_\pi^{-1}\left\{\limsup_{n\to\infty} S_n \leq x\right\} = \left\{\limsup_{n\to\infty} S_n\circ T_\pi \leq x\right\} = \left\{\limsup_{n\to\infty} S_n \leq x\right\}$$
which shows that $\limsup_{n\to\infty} S_n$ is $\mathcal{S}$ – measurable. Therefore, $\limsup_{n\to\infty} S_n = c$ a.s. for some $c \in \bar{\mathbb{R}}$. Since $(X_2, X_3, \dots) \overset{d}{=} (X_1, X_2, \dots)$ it follows (see Corollary 6.47 and Exercise 6.10) that
$$c = \limsup_{n\to\infty} S_n \overset{d}{=} \limsup_{n\to\infty}(X_2 + X_3 + \dots + X_{n+1})$$
$$= \limsup_{n\to\infty}(S_{n+1} - X_1) = \limsup_{n\to\infty} S_{n+1} - X_1 = c - X_1.$$

By Exercise 10.7 below we may now conclude that $c = c - X_1$ a.s. which is possible iff $c \in \{\pm\infty\}$ or $X_1 = 0$ a.s. Since the latter is not allowed, $\limsup_{n\to\infty} S_n = \infty$ or $\limsup_{n\to\infty} S_n = -\infty$ a.s.
3. Now assume that $P(X_1 \neq 0) > 0$ and $X_1 \overset{\mathrm{d}}{=} -X_1$, i.e. $P(X_1 \in A) = P(-X_1 \in A)$ for all $A \in \mathcal{B}_{\mathbb{R}}$. By 2. we know $\limsup_{n\to\infty} S_n = c$ a.s. with $c \in \{\pm\infty\}$. Since $\{X_n\}_{n=1}^\infty$ and $\{-X_n\}_{n=1}^\infty$ are i.i.d. and $-X_n \overset{\mathrm{d}}{=} X_n$, it follows that $\{X_n\}_{n=1}^\infty \overset{\mathrm{d}}{=} \{-X_n\}_{n=1}^\infty$ .The results of Exercises 6.10 and 10.7 then imply that $c \overset{\mathrm{d}}{=} \limsup_{n\to\infty} S_n \overset{\mathrm{d}}{=} \limsup_{n\to\infty}(-S_n)$ and in particular
$$c \overset{\text{a.s.}}{=} \limsup_{n\to\infty}(-S_n) = -\liminf_{n\to\infty} S_n \geq -\limsup_{n\to\infty} S_n = -c.$$
Since the $c = -\infty$ does not satisfy, $c \geq -c$, we must $c = \infty$. Hence in this symmetric case we have shown,
$$\limsup_{n\to\infty} S_n = \infty \text{ and } \liminf_{n\to\infty} S_n = -\infty \text{ a.s.}$$

Exercise 10.7. Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $Y : \Omega \to \bar{\mathbb{R}}$ is a random variable and $c \in \bar{\mathbb{R}}$ is a constant. Then $Y = c$ a.s. iff $Y \overset{d}{=} c$.

Solution to Exercise (10.7). If $Y = c$ a.s. then $P(Y \in A) = P(c \in A)$ for all $A \in \mathcal{B}_{\mathbb{R}}$ and therefore $Y \overset{d}{=} c$. Conversely, if $Y \overset{d}{=} c$, then $P(Y = c) = P(c = c) = 1$, i.e. $Y = c$ a.s.

10.9 Another Construction of Independent Random Variables*

This section may be skipped as the results are a special case of those given above. The arguments given here avoid the use of Kolmogorov's existence theorem for product measures.

Example 10.52. Suppose that $\Omega = \Lambda^n$ where Λ is a finite set, $\mathcal{B} = 2^\Omega$, $P(\{\omega\}) = \prod_{j=1}^n q_j(\omega_j)$ where $q_j : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_j(\lambda) = 1$. Let $\mathcal{C}_i := \left\{\Lambda^{i-1}\times A \times \Lambda^{n-i} : A \subset \Lambda\right\}$. Then $\{\mathcal{C}_i\}_{i=1}^n$ are independent. Indeed, if $B_i := \Lambda^{i-1} \times A_i \times \Lambda^{n-i}$, then

$$\cap B_i = A_1 \times A_2 \times \dots \times A_n$$

and we have

$$P(\cap B_i) = \sum_{\omega \in A_1 \times A_2 \times \cdots \times A_n} \prod_{i=1}^n q_i(\omega_i) = \prod_{i=1}^n \sum_{\lambda \in A_i} q_i(\lambda)$$

while

$$P(B_i) = \sum_{\omega \in \Lambda^{i-1} \times A_i \times \Lambda^{n-i}} \prod_{i=1}^n q_i(\omega_i) = \sum_{\lambda \in A_i} q_i(\lambda).$$

Example 10.53. Continue the notation of Example 10.52 and further assume that $\Lambda \subset \mathbb{R}$ and let $X_i : \Omega \to \Lambda$ be defined by, $X_i(\omega) = \omega_i$. Then $\{X_i\}_{i=1}^n$ are independent random variables. Indeed, $\sigma(X_i) = \mathcal{C}_i$ with $\mathcal{C}_i$ as in Example 10.52.

Alternatively, from Exercise 4.10, we know that

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

for all $f_i : \Lambda \to \mathbb{R}$. Taking $A_i \subset \Lambda$ and $f_i := 1_{A_i}$ in the above identity shows that

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{E}_P\left[\prod_{i=1}^n 1_{A_i}(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[1_{A_i}(X_i)]$$
$$= \prod_{i=1}^n P(X_i \in A_i)$$

as desired.

Theorem 10.54 (Existence of i.i.d simple R.V.'s). *Suppose that $\{q_i\}_{i=0}^n$ is a sequence of positive numbers such that $\sum_{i=0}^n q_i = 1$. Then there exists a sequence $\{X_k\}_{k=1}^\infty$ of simple random variables taking values in $\Lambda = \{0, 1, 2 \dots, n\}$ on $((0,1], \mathcal{B}, m)$ such that*

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}$$

for all $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$ and all $k \in \mathbb{N}$. (See Example 10.15 above and Theorem 10.58 below for the general case of this theorem.)

Proof. For $i = 0, 1, \dots, n$, let $\sigma_{-1} = 0$ and $\sigma_j := \sum_{i=0}^j q_i$ and for any interval, $(a, b]$, let

$$T_i((a,b]) := (a + \sigma_{i-1}(b-a), a + \sigma_i(b-a)].$$

Given $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$, let

$$J_{i_1, i_2, \dots, i_k} := T_{i_k}\left(T_{i_{k-1}}(\dots T_{i_1}((0,1]))\right)$$

and define $\{X_k\}_{k=1}^\infty$ on $(0,1]$ by

$$X_k := \sum_{i_1, i_2, \dots, i_k \in \{0,1,2,\dots,n\}} i_k 1_{J_{i_1, i_2, \dots, i_k}},$$

see Figure 10.3. Repeated applications of Corollary 6.27 shows the functions, $X_k : (0,1] \to \mathbb{R}$ are measurable.

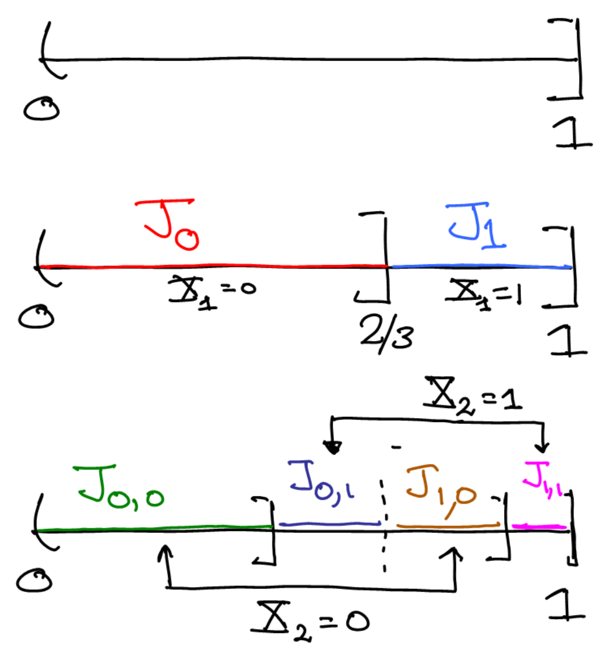


Fig. 10.3. Here we suppose that $p_0 = 2/3$ and $p_1 = 1/3$ and then we construct J_l and $J_{l,k}$ for $l, k \in \{0, 1\}$.

Observe that

$$m(T_i((a,b])) = q_i(b-a) = q_i m((a,b]), \tag{10.37}$$

and so by induction,

$$m(J_{i_1, i_2, \dots, i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}.$$

The reader should convince herself/himself that

$$\{X_1 = i_1, \dots X_k = i_i\} = J_{i_1, i_2, \dots, i_k}$$

and therefore, we have

$$m(\{X_1 = i_1, \dots, X_k = i_i\}) = m(J_{i_1,i_2,\dots,i_k}) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}$$

as desired. ■

Corollary 10.55 (Independent variables on product spaces). *Suppose $\Lambda = \{0, 1, 2 \dots, n\}$, $q_i > 0$ with $\sum_{i=0}^n q_i = 1$, $\Omega = \Lambda^\infty = \Lambda^{\mathbb{N}}$, and for $i \in \mathbb{N}$, let $Y_i : \Omega \to \mathbb{R}$ be defined by $Y_i(\omega) = \omega_i$ for all $\omega \in \Omega$. Further let $\mathcal{B} := \sigma(Y_1, Y_2, \dots, Y_n, \dots)$. Then there exists a unique probability measure, $P : \mathcal{B} \to [0, 1]$ such that*

$$P(\{Y_1 = i_1, \dots, Y_k = i_i\}) = q_{i_1} \dots q_{i_k}.$$

Proof. Let $\{X_i\}_{i=1}^n$ be as in Theorem 10.54 and define $T : (0, 1] \to \Omega$ by

$$T(x) = (X_1(x), X_2(x), \dots, X_k(x), \dots).$$

Observe that T is measurable since $Y_i \circ T = X_i$ is measurable for all i. We now define, $P := T_* m$. Then we have

$$\begin{aligned} P(\{Y_1 = i_1, \dots, Y_k = i_i\}) &= m\left(T^{-1}(\{Y_1 = i_1, \dots, Y_k = i_i\})\right) \\ &= m(\{Y_1 \circ T = i_1, \dots, Y_k \circ T = i_i\}) \\ &= m(\{X_1 = i_1, \dots, X_k = i_i\}) = q_{i_1} \dots q_{i_k}. \end{aligned}$$

■

Theorem 10.56. *Given a finite subset, $\Lambda \subset \mathbb{R}$ and a function $q : \Lambda \to [0, 1]$ such that $\sum_{\lambda \in \Lambda} q(\lambda) = 1$, there exists a probability space, $(\Omega, \mathcal{B}, P)$ and an independent sequence of random variables, $\{X_n\}_{n=1}^\infty$ such that $P(X_n = \lambda) = q(\lambda)$ for all $\lambda \in \Lambda$.*

Proof. Use Corollary 10.10 to shows that random variables constructed in Example 5.41 or Theorem 10.54 fit the bill. ■

Proposition 10.57. *Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of i.i.d. random variables with distribution, $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$. If we let $U := \sum_{n=1}^\infty 2^{-n} X_n$, then $P(U \le x) = (0 \vee x) \wedge 1$, i.e. U has the uniform distribution on $[0, 1]$.*

Proof. Let us recall that $P(X_n = 0 \text{ a.a.}) = 0 = P(X_n = 1 \text{ a.a.})$. Hence we may, by shrinking Ω if necessary, assume that $\{X_n = 0 \text{ a.a.}\} = \emptyset = \{X_n = 1 \text{ a.a.}\}$. With this simplification, we have

$$\begin{aligned} \left\{U < \frac{1}{2}\right\} &= \{X_1 = 0\}, \\ \left\{U < \frac{1}{4}\right\} &= \{X_1 = 0, X_2 = 0\} \text{ and} \\ \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} &= \{X_1 = 1, X_2 = 0\} \end{aligned}$$

and hence that

$$\begin{aligned} \left\{U < \frac{3}{4}\right\} &= \left\{U < \frac{1}{2}\right\} \cup \left\{\frac{1}{2} \le U < \frac{3}{4}\right\} \\ &= \{X_1 = 0\} \cup \{X_1 = 1, X_2 = 0\}. \end{aligned}$$

From these identities, it follows that

$$P(U < 0) = 0,\ P\left(U < \frac{1}{4}\right) = \frac{1}{4},\ P\left(U < \frac{1}{2}\right) = \frac{1}{2},\ \text{and } P\left(U < \frac{3}{4}\right) = \frac{3}{4}.$$

More generally, we claim that if $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ with $\varepsilon_j \in \{0, 1\}$, then

$$P(U < x) = x. \tag{10.38}$$

The proof is by induction on n. Indeed, we have already verified (10.38) when $n = 1, 2$. Suppose we have verified (10.38) up to some $n \in \mathbb{N}$ and let $x = \sum_{j=1}^n \varepsilon_j 2^{-j}$ and consider

$$\begin{aligned} P\left(U < x + 2^{-(n+1)}\right) &= P(U < x) + P\left(x \le U < x + 2^{-(n+1)}\right) \\ &= x + P\left(x \le U < x + 2^{-(n+1)}\right). \end{aligned}$$

Since

$$\left\{x \le U < x + 2^{-(n+1)}\right\} = \left[\cap_{j=1}^n \{X_j = \varepsilon_j\}\right] \cap \{X_{n+1} = 0\}$$

we see that

$$P\left(x \le U < x + 2^{-(n+1)}\right) = 2^{-(n+1)}$$

and hence

$$P\left(U < x + 2^{-(n+1)}\right) = x + 2^{-(n+1)}$$

which completes the induction argument.

Since $x \to P(U < x)$ is left continuous we may now conclude that $P(U < x) = x$ for all $x \in (0, 1)$ and since $x \to x$ is continuous we may also deduce that $P(U \le x) = x$ for all $x \in (0, 1)$. Hence we may conclude that

$$P(U \le x) = (0 \vee x) \wedge 1.$$

■

We may now show the existence of independent random variables with arbitrary distributions.

Theorem 10.58. *Suppose that* $\{\mu_n\}_{n=1}^{\infty}$ *are a sequence of probability measures on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. *Then there exists a probability space,* $(\Omega, \mathcal{B}, P)$ *and a sequence* $\{Y_n\}_{n=1}^{\infty}$ *independent random variables with* $\mathrm{Law}(Y_n) := P \circ Y_n^{-1} = \mu_n$ *for all* n.

Proof. By Theorem 10.56, there exists a sequence of i.i.d. random variables, $\{Z_n\}_{n=1}^{\infty}$, such that $P(Z_n = 1) = P(Z_n = 0) = \frac{1}{2}$. These random variables may be put into a two dimensional array, $\{X_{i,j} : i, j \in \mathbb{N}\}$, see the proof of Lemma 3.8. For each i, let $U_i := \sum_{j=1}^{\infty} 2^{-i} X_{i,j}$ – $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable random variable. According to Proposition 10.57, U_i is uniformly distributed on $[0, 1]$. Moreover by the grouping Lemma 10.16, $\left\{\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)\right\}_{i=1}^{\infty}$ are independent σ – algebras and hence $\{U_i\}_{i=1}^{\infty}$ is a sequence of i.i.d.. random variables with the uniform distribution.

Finally, let $F_i(x) := \mu((-\infty, x])$ for all $x \in \mathbb{R}$ and let $G_i(y) = \inf\{x : F_i(x) \geq y\}$. Then according to Theorem 6.48, $Y_i := G_i(U_i)$ has μ_i as its distribution. Moreover each Y_i is $\sigma\left(\{X_{i,j}\}_{j=1}^{\infty}\right)$ – measurable and therefore the $\{Y_i\}_{i=1}^{\infty}$ are independent random variables. ■

The Standard Poisson Process

11.1 Poisson Random Variables

Recall from Exercise 7.5 that a Random variable, X, is Poisson distributed with intensity, a, if

$$P(X=k)=\frac{a^k}{k!}e^{-a} \text{ for all } k\in\mathbb{N}_0.$$

We will abbreviate this in the future by writing $X\overset{d}{=}\mathrm{Poi}(a)$. Let us also recall that

$$\mathbb{E}\left[z^X\right]=\sum_{k=0}^{\infty}z^k\frac{a^k}{k!}e^{-a}=e^{az}e^{-a}=e^{a(z-1)}$$

and as in Exercise 7.5 we have $\mathbb{E}X=a=\mathrm{Var}(X)$.

Lemma 11.1. *If $X=\mathrm{Poi}(a)$ and $Y=\mathrm{Poi}(b)$ and X and Y are independent, then $X+Y=\mathrm{Poi}(a+b)$.*

Proof. For $k\in\mathbb{N}_0$,

$$\begin{aligned}P(X+Y=k)&=\sum_{l=0}^{k}P(X=l,Y=k-l)=\sum_{l=0}^{k}P(X=l)P(Y=k-l)\\&=\sum_{l=0}^{k}e^{-a}\frac{a^l}{l!}e^{-b}\frac{b^{k-l}}{(k-l)!}=\frac{e^{-(a+b)}}{k!}\sum_{l=0}^{k}\binom{k}{l}a^lb^{k-l}\\&=\frac{e^{-(a+b)}}{k!}(a+b)^k.\end{aligned}$$

Alternative Proof. Notice that

$$\mathbb{E}\left[z^{X+Y}\right]=\mathbb{E}\left[z^X\right]\mathbb{E}\left[z^Y\right]=e^{a(z-1)}e^{b(z-1)}=\exp\left((a+b)(z-1)\right).$$

This suffices to complete the proof. ■

Lemma 11.2. *Suppose that $\{N_i\}_{i=1}^\infty$ are independent Poisson random variables with parameters, $\{\lambda_i\}_{i=1}^\infty$ such that $\sum_{i=1}^\infty\lambda_i=\infty$. Then $\sum_{i=1}^\infty N_i=\infty$ a.s.*

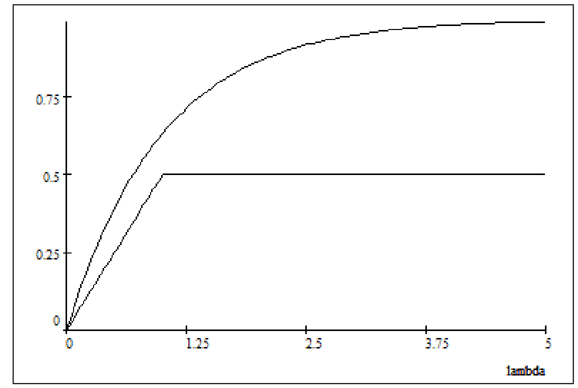


Fig. 11.1. This plot shows, $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$.

Proof. From Figure 11.1 we see that $1-e^{-\lambda}\geq\frac{1}{2}(1\wedge\lambda)$ for all $\lambda\geq 0$. Therefore,

$$\sum_{i=1}^\infty P(N_i\geq 1)=\sum_{i=1}^\infty\left(1-P(N_i=0)\right)=\sum_{i=1}^\infty\left(1-e^{-\lambda_i}\right)\geq\frac{1}{2}\sum_{i=1}^\infty\lambda_i\wedge 1=\infty$$

and so by the first Borel Cantelli Lemma, $P(\{N_i\geq 1\text{ i.o.}\})=1$. From this it certainly follows that $\sum_{i=1}^\infty N_i=\infty$ a.s.

Alternatively, let $\Lambda_n=\lambda_1+\dots+\lambda_n$, then

$$P\left(\sum_{i=1}^\infty N_i\geq k\right)\geq P\left(\sum_{i=1}^n N_i\geq k\right)=1-e^{-\Lambda_n}\sum_{l=0}^{k-1}\frac{\Lambda_n^l}{l!}\to 1\text{ as }n\to\infty.$$

Therefore $P\left(\sum_{i=1}^\infty N_i\geq k\right)=1$ for all $k\in\mathbb{N}$ and hence,

$$P\left(\sum_{i=1}^\infty N_i\geq\infty\right)=P\left(\cap_{k=1}^\infty\left\{\sum_{i=1}^\infty N_i\geq k\right\}\right)=1.$$

■

11.2 Exponential Random Variables

Recall from Definition 7.55 that $T \stackrel{d}{=} E(\lambda)$ is an exponential random variable with parameter $\lambda \in [0, \infty)$ provided, $P(T > t) = e^{-\lambda t}$ for all $t \geq 0$. We have seen that

$$\mathbb{E}\left[e^{aT}\right] = \sum_{n=0}^{\infty} \frac{a^n}{n!} \mathbb{E}[T^n] = \frac{1}{1 - a\lambda^{-1}} \text{ for } |a| < \lambda. \tag{11.1}$$

$\mathbb{E}T = \lambda^{-1}$ and $\operatorname{Var}(T) = \lambda^{-2}$, and (see Theorem 7.56) that T being exponential is characterized by the following memoryless property;

$$P(T > s + t | T > s) = P(T > t) \text{ for all } s, t \geq 0.$$

Theorem 11.3. *Let $\{T_j\}_{j=1}^{\infty}$ be independent random variables such that $T_j \stackrel{d}{=} E(\lambda_j)$ with $0 < \lambda_j < \infty$ for all j. Then:*

1. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$ then $P(\sum_{n=1}^{\infty} T_n = \infty) = 0$ (i.e. $P(\sum_{n=1}^{\infty} T_n < \infty) = 1$).*
2. *If $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$ then $P(\sum_{n=1}^{\infty} T_n = \infty) = 1$.*

(By Kolmogorov's zero-one law (see Proposition 10.46) it follows that $P(\sum_{n=1}^{\infty} T_n = \infty)$ is always either 0 or 1. We are showing here that $P(\sum_{n=1}^{\infty} T_n = \infty) = 1$ iff $\mathbb{E}[\sum_{n=1}^{\infty} T_n] = \infty$.)

Proof. 1. Since

$$\mathbb{E}\left[\sum_{n=1}^{\infty} T_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[T_n] = \sum_{n=1}^{\infty} \lambda_n^{-1} < \infty$$

it follows that $\sum_{n=1}^{\infty} T_n < \infty$ a.s., i.e. $P(\sum_{n=1}^{\infty} T_n = \infty) = 0$.

2. By the DCT, independence, and Eq. (11.1) with $a = 1$,

$$\begin{aligned} \mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] &= \lim_{N \to \infty} \mathbb{E}\left[e^{-\sum_{n=1}^{N} T_n}\right] = \lim_{N \to \infty} \prod_{n=1}^{N} \mathbb{E}\left[e^{-T_n}\right] \\ &= \lim_{N \to \infty} \prod_{n=1}^{N} \left(\frac{1}{1 + \lambda_n^{-1}}\right) = \prod_{n=1}^{\infty} (1 - a_n) \end{aligned}$$

where

$$a_n = 1 - \frac{1}{1 + \lambda_n^{-1}} = \frac{1}{1 + \lambda_n}.$$

Hence by Exercise 10.6, $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $\infty = \sum_{n=1}^{\infty} a_n$ which happens iff $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$. This completes the proof since $\mathbb{E}\left[e^{-\sum_{n=1}^{\infty} T_n}\right] = 0$ iff $e^{-\sum_{n=1}^{\infty} T_n} = 0$ a.s. or equivalently $\sum_{n=1}^{\infty} T_n = \infty$ a.s. ■

11.2.1 Appendix: More properties of Exponential random Variables*

Theorem 11.4. *Let I be a countable set and let $\{T_k\}_{k \in I}$ be independent random variables such that $T_k \sim E(q_k)$ with $q := \sum_{k \in I} q_k \in (0, \infty)$. Let $T := \inf_k T_k$ and let $K = k$ on the set where $T_j > T_k$ for all $j \neq k$. On the complement of all these sets, define $K = *$ where $*$ is some point not in I. Then $P(K = *) = 0$, K and T are independent, $T \sim E(q)$, and $P(K = k) = q_k / q$.*

Proof. Let $k \in I$ and $t \in \mathbb{R}_+$ and $\Lambda_n \subset_f I$ such that $\Lambda_n \uparrow I \setminus \{k\}$, then

$$\begin{aligned} P(K = k, T > t) &= P(\cap_{j \neq k} \{T_j > T_k\}, T_k > t) = \lim_{n \to \infty} P(\cap_{j \in \Lambda_n} \{T_j > T_k\}, T_k > t) \\ &= \lim_{n \to \infty} \int_{[0,\infty)^{\Lambda_n \cup \{k\}}} \prod_{j \in \Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n \left(\{t_j\}_{j \in \Lambda_n}\right) q_k e^{-q_k t_k} dt_k \end{aligned}$$

where μ_n is the joint distribution of $\{T_j\}_{j \in \Lambda_n}$. So by Fubini's theorem,

$$\begin{aligned} P(K = k, T > t) &= \lim_{n \to \infty} \int_t^{\infty} q_k e^{-q_k t_k} dt_k \int_{[0,\infty)^{\Lambda_n}} \prod_{j \in \Lambda_n} 1_{t_j > t_k} \cdot 1_{t_k > t} d\mu_n \left(\{t_j\}_{j \in \Lambda_n}\right) \\ &= \lim_{n \to \infty} \int_t^{\infty} P(\cap_{j \in \Lambda_n} \{T_j > t_k\}) q_k e^{-q_k t_k} dt_k \\ &= \int_t^{\infty} P(\cap_{j \neq k} \{T_j > \tau\}) q_k e^{-q_k \tau} d\tau \\ &= \int_t^{\infty} \prod_{j \neq k} e^{-q_j \tau} q_k e^{-q_k \tau} d\tau = \int_t^{\infty} \prod_{j \in I} e^{-q_j \tau} q_k d\tau \\ &= \int_t^{\infty} e^{-\sum_{j=1}^{\infty} q_j \tau} q_k d\tau = \int_t^{\infty} e^{-q\tau} q_k d\tau = \frac{q_k}{q} e^{-qt}. \end{aligned} \tag{11.2}$$

Taking $t = 0$ shows that $P(K = k) = \frac{q_k}{q}$ and summing this on k shows $P(K \in I) = 1$ so that $P(K = *) = 0$. Moreover summing Eq. (11.2) on k now shows that $P(T > t) = e^{-qt}$ so that T is exponential. Moreover we have shown that

$$P(K = k, T > t) = P(K = k) P(T > t)$$

proving the desired independence. ■

Theorem 11.5. *Suppose that $S \sim E(\lambda)$ and $R \sim E(\mu)$ are independent. Then for $t \geq 0$ we have*

$$\mu P(S \leq t < S + R) = \lambda P(R \leq t < R + S).$$

Proof. We have

$$\begin{aligned}\mu P(S\le t<S+R) &= \mu\int_0^t \lambda e^{-\lambda s}P(t<s+R)\,ds\\ &= \mu\lambda\int_0^t e^{-\lambda s}e^{-\mu(t-s)}ds\\ &= \mu\lambda e^{-\mu t}\int_0^t e^{-(\lambda-\mu)s}ds = \mu\lambda e^{-\mu t}\cdot\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\ &= \mu\lambda\cdot\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}\end{aligned}$$

which is symmetric in the interchanged of μ and λ.**Alternatively:**

$$\begin{aligned}P(S\le t<S+R) &= \lambda\mu\int_{\mathbb{R}_+^2} 1_{s\le t<s+r}e^{-\lambda s}e^{-\mu r}dsdr\\ &= \lambda\mu\int_0^t ds\int_{t-s}^\infty dre^{-\lambda s}e^{-\mu r}\\ &= \lambda\int_0^t dse^{-\lambda s}e^{-\mu(t-s)}\\ &= \lambda e^{-\mu t}\int_0^t dse^{-(\lambda-\mu)s}\\ &= \lambda e^{-\mu t}\frac{1-e^{-(\lambda-\mu)t}}{\lambda-\mu}\\ &= \lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.\end{aligned}$$

Therefore,

$$\mu P(S\le t<S+R) = \mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}$$

which is symmetric in the interchanged of μ and λ and hence

$$\lambda P(R\le t<S+R) = \mu\lambda\frac{e^{-\mu t}-e^{-\lambda t}}{\lambda-\mu}.$$

∎

Example 11.6. Suppose T is a positive random variable such that $P(T\ge t+s|T\ge s) = P(T\ge t)$ for all $s,t\ge 0$, or equivalently

$$P(T\ge t+s) = P(T\ge t)\,P(T\ge s)\ \text{ for all } s,t\ge 0,$$

then $P(T\ge t) = e^{-at}$ for some $a>0$. (Such exponential random variables are often used to model "waiting times.") The distribution function for T is $F_T(t) := P(T\le t) = 1-e^{-a(t\vee 0)}$. Since $F_T(t)$ is piecewise differentiable, the law of T, $\mu := P\circ T^{-1}$, has a density,

$$d\mu(t) = F_T'(t)\,dt = ae^{-at}1_{t\ge 0}dt.$$

Therefore,

$$\mathbb{E}\left[e^{iaT}\right] = \int_0^\infty ae^{-at}e^{i\lambda t}dt = \frac{a}{a-i\lambda} = \hat\mu(\lambda).$$

Since

$$\hat\mu'(\lambda) = i\frac{a}{(a-i\lambda)^2}\ \text{and}\ \hat\mu''(\lambda) = -2\frac{a}{(a-i\lambda)^3}$$

it follows that

$$\mathbb{E}T = \frac{\hat\mu'(0)}{i} = a^{-1}\ \text{and}\ \mathbb{E}T^2 = \frac{\hat\mu''(0)}{i^2} = \frac{2}{a^2}$$

and hence $\mathrm{Var}(T) = \frac{2}{a^2}-\left(\frac{1}{a}\right)^2 = a^{-2}$.

11.3 The Standard Poisson Process

Let $\{T_k\}_{k=1}^\infty$ be an i.i.d. sequence of random exponential times with parameter λ, i.e. $P(T_k\in[t,t+dt]) = \lambda e^{-\lambda t}dt$. For each $n\in\mathbb{N}$ let $W_n := T_1+\dots+T_n$ be the "**waiting time**" for the n^{th} event to occur. Because of Theorem 11.3 we know that $\lim_{n\to\infty}W_n = \infty$ a.s.

Definition 11.7 (Poisson Process I). *For any subset $A\subset\mathbb{R}_+$ let $N(A) := \sum_{n=1}^\infty 1_A(W_n)$ count the number of waiting times which occurred in A. When $A = (0,t]$ we will write, $N_t := N((0,t])$ for all $t\ge 0$ and refer to $\{N_t\}_{t\ge 0}$ as the **Poisson Process with intensity** λ. (Observe that $\{N_t = n\} = W_n\le t< W_{n+1}$.)*

The next few results summarize a number of the basic properties of this Poisson process. Many of the proofs will be left as exercises to the reader. We will use the following notation below; for each $n\in\mathbb{N}$ and $T\ge 0$ let

$$\Delta_n(T) := \{(w_1,\dots,w_n)\in\mathbb{R}^n : 0<w_1<w_2<\dots<w_n<T\}$$

and let

$$\Delta_n := \cup_{T>0}\Delta_n(T) = \{(w_1,\dots,w_n)\in\mathbb{R}^n : 0<w_1<w_2<\dots<w_n<\infty\}.$$

Exercise 11.1. Show $m_n(\Delta_n(T)) = T^n/n!$ where m_n is Lebesgue measure on $\mathcal{B}_{\mathbb{R}^n}$.

Exercise 11.2. If $n \in \mathbb{N}$ and $g : \Delta_n \to \mathbb{R}$ bounded (non-negative) measurable, then

$$\mathbb{E}[g(W_1, \ldots, W_n)] = \int_{\Delta_n} g(w_1, w_2, \ldots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \ldots dw_n. \tag{11.3}$$

As a simple corollary we have the following direct proof of Example 10.28.

Corollary 11.8. *If $n \in \mathbb{N}$, then $W_n \stackrel{d}{=} \text{Gamma}(n, \lambda^{-1})$.*

Proof. Taking $g(w_1, w_2, \ldots, w_n) = g(w_n)$ in Eq. (11.3) we find with the aid of Exercise 11.1 that

$$\begin{aligned}\mathbb{E}[f(W_n)] &= \int_{\Delta_n} f(w_n) \lambda^n e^{-\lambda w_n} dw_1 \ldots dw_n \\ &= \int_0^\infty f(w) \lambda^n \frac{w^{n-1}}{(n-1)!} e^{-\lambda w} dw\end{aligned}$$

which shows that $W_n \stackrel{d}{=} \text{Gamma}(n, \lambda^{-1})$. ■

Corollary 11.9. *If $t \in \mathbb{R}_+$ and $f : \Delta_n(t) \to \mathbb{R}$ is a bounded (or non-negative) measurable function, then*

$$\begin{aligned}&\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n] \\ &\quad = \lambda^n e^{-\lambda t} \int_{\Delta_n(t)} f(w_1, w_2, \ldots, w_n) dw_1 \ldots dw_n.\end{aligned} \tag{11.4}$$

Proof. Making use of the observation that $\{N_t = n\} = \{W_n \le t < W_{n+1}\}$, we may apply Eq. (11.3) at level $n+1$ with

$$g(w_1, w_2, \ldots, w_{n+1}) = f(w_1, w_2, \ldots, w_n) 1_{w_n \le t < w_{n+1}}$$

to learn

$$\begin{aligned}&\mathbb{E}[f(W_1, \ldots, W_n) : N_t = n] \\ &= \int_{0 < w_1 < \cdots < w_n < t < w_{n+1}} f(w_1, w_2, \ldots, w_n) \lambda^{n+1} e^{-\lambda w_{n+1}} dw_1 \ldots dw_n dw_{n+1} \\ &= \int_{\Delta_n(t)} f(w_1, w_2, \ldots, w_n) \lambda^n e^{-\lambda t} dw_1 \ldots dw_n.\end{aligned}$$

■

Exercise 11.3. Show $N_t \stackrel{d}{=} \text{Poi}(\lambda t)$ for all $t > 0$.

Definition 11.10 (Order Statistics). *Suppose that $X_1, \ldots, X_n$ are non-negative random variables such that $P(X_i = X_j) = 0$ for all $i \neq j$. The **order statistics** of $X_1, \ldots, X_n$ is a collection of random variables, $\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n$ such that $\{X_1, \ldots, X_n\} = \left\{\tilde{X}_1 < \tilde{X}_2 < \cdots < \tilde{X}_n\right\}$. (The order statistics are uniquely determined off the set $\cup_{i \neq j}\{X_i = X_j\}$ which is a null set.)*

Exercise 11.4. Suppose that $X_1, \ldots, X_n$ are non-negative[1] random variables such that $P(X_i = X_j) = 0$ for all $i \neq j$. Show;

1. If $f : \Delta_n \to \mathbb{R}$ is bounded (non-negative) measurable, then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = \sum_{\sigma \in S_n} \mathbb{E}[f(X_{\sigma 1}, \ldots, X_{\sigma n}) : X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}], \tag{11.5}$$
where S_n is the permutation group on $\{1, 2, \ldots, n\}$.
2. If we further assume that $\{X_1, \ldots, X_n\}$ are i.i.d. random variables, then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = n! \cdot \mathbb{E}[f(X_1, \ldots, X_n) : X_1 < X_2 < \cdots < X_n]. \tag{11.6}$$
3. $f : \mathbb{R}_+^n \to \mathbb{R}$ is a bounded (non-negative) measurable symmetric function (i.e. $f(w_{\sigma 1}, \ldots, w_{\sigma n}) = f(w_1, \ldots, w_n)$ for all $\sigma \in S_n$ and $(w_1, \ldots, w_n) \in \mathbb{R}_+^n$) then
$$\mathbb{E}\left[f\left(\tilde{X}_1, \ldots, \tilde{X}_n\right)\right] = \mathbb{E}[f(X_1, \ldots, X_n)].$$
4. Suppose that $Y_1, \ldots, Y_n$ is another collection of non-negative random variables such that $P(Y_i = Y_j) = 0$ for all $i \neq j$ such that
$$\mathbb{E}[f(X_1, \ldots, X_n)] = \mathbb{E}[f(Y_1, \ldots, Y_n)]$$
for all bounded (non-negative) measurable symmetric functions from $\mathbb{R}_+^n \to \mathbb{R}$. Show that $\left(\tilde{X}_1, \ldots, \tilde{X}_n\right) \stackrel{d}{=} \left(\tilde{Y}_1, \ldots, \tilde{Y}_n\right)$.
Hint: if $g : \Delta_n \to \mathbb{R}$ is a bounded measurable function, define $f : \mathbb{R}_+^n \to \mathbb{R}$ by;
$$f(y_1, \ldots, y_n) = \sum_{\sigma \in S_n} 1_{y_{\sigma 1} < y_{\sigma 2} < \cdots < y_{\sigma n}} g(y_{\sigma 1}, y_{\sigma 2}, \ldots, y_{\sigma n})$$
and then show f is symmetric.

[1] The non-negativity of the X_i are not really necessary here but this is all we need to consider.

Exercise 11.5. Let $t \in \mathbb{R}_+$ and $\{U_i\}_{i=1}^n$ be i.i.d. uniformly distributed random variables on $[0,t]$. Show that the order statistics, $\left(\tilde{U}_1, \dots, \tilde{U}_n\right)$, of $(U_1, \dots, U_n)$ has the same distribution as $(W_1, \dots, W_n)$ given $N_t = n$. (Thus, given $N_t = n$, the collection of points, $\{W_1, \dots, W_n\}$, has the same distribution as the collection of points, $\{U_1, \dots, U_n\}$, in $[0,t]$.)

Theorem 11.11 (Joint Distributions). *If $\{A_i\}_{i=1}^k \subset \mathcal{B}_{[0,t]}$ is a partition of $[0,t]$, then $\{N(A_i)\}_{i=1}^k$ are independent random variables and $N(A) \overset{d}{=} \mathrm{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_{[0,t]}$ with $m(A) < \infty$. In particular, if $0 < t_1 < t_2 < \dots < t_n$, then $\left\{N_{t_i} - N_{t_{i-1}}\right\}_{i=1}^n$ are independent random variables and $N_t - N_s \overset{d}{=} \mathrm{Poi}(\lambda(t-s))$ for all $0 \le s < t < \infty$. (We say that $\{N_t\}_{t\ge 0}$ is a stochastic process with **independent increments**.)*

Proof. If $z \in \mathbb{C}$ and $A \in \mathcal{B}_{[0,t]}$, then

$$z^{N(A)} = z^{\sum_{i=1}^n 1_A(W_i)} \text{ on } \{N_t = n\}.$$

Let $n \in \mathbb{N}$, $z_i \in \mathbb{C}$, and define

$$f(w_1, \dots, w_n) = z_1^{\sum_{i=1}^n 1_{A_1}(w_i)} \dots z_k^{\sum_{i=1}^n 1_{A_k}(w_i)}$$

which is a symmetric function. On $N_t = n$ we have,

$$z_1^{N(A_1)} \dots z_k^{N(A_k)} = f(W_1, \dots, W_n)$$

and therefore,

$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \dots z_k^{N(A_k)} | N_t = n\right] &= \mathbb{E}\left[f(W_1, \dots, W_n) | N_t = n\right] \\
&= \mathbb{E}\left[f(U_1, \dots, U_n)\right] \\
&= \mathbb{E}\left[z_1^{\sum_{i=1}^n 1_{A_1}(U_i)} \dots z_k^{\sum_{i=1}^n 1_{A_k}(U_i)}\right] \\
&= \prod_{i=1}^n \mathbb{E}\left[\left(z_1^{1_{A_1}(U_i)} \dots z_k^{1_{A_k}(U_i)}\right)\right] \\
&= \left(\mathbb{E}\left[\left(z_1^{1_{A_1}(U_1)} \dots z_k^{1_{A_k}(U_1)}\right)\right]\right)^n \\
&= \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n,
\end{aligned}$$

wherein we have made use of the fact that $\{A_i\}_{i=1}^n$ is a partition of $[0,t]$ so that

$$z_1^{1_{A_1}(U_1)} \dots z_k^{1_{A_k}(U_1)} = \sum_{i=1}^k z_i 1_{A_i}(U_i).$$

Thus it follows that

$$\begin{aligned}
\mathbb{E}\left[z_1^{N(A_1)} \dots z_k^{N(A_k)}\right] &= \sum_{n=0}^\infty \mathbb{E}\left[z_1^{N(A_1)} \dots z_k^{N(A_k)} | N_t = n\right] P(N_t = n) \\
&= \sum_{n=0}^\infty \left(\frac{1}{t}\sum_{i=1}^k m(A_i) \cdot z_i\right)^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} \\
&= \sum_{n=0}^\infty \frac{1}{n!}\left(\lambda \sum_{i=1}^k m(A_i) \cdot z_i\right)^n e^{-\lambda t} \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i) z_i - t\right]\right) \\
&= \exp\left(\lambda\left[\sum_{i=1}^k m(A_i)(z_i - 1)\right]\right).
\end{aligned}$$

From this result it follows that $\{N(A_i)\}_{i=1}^n$ are independent random variables and $N(A) = \mathrm{Poi}(\lambda m(A))$ for all $A \in \mathcal{B}_{\mathbb{R}}$ with $m(A) < \infty$.

Alternatively; suppose that $a_i \in \mathbb{N}_0$ and $n := a_1 + \dots + a_k$, then

$$\begin{aligned}
P[N(A_1) = a_1, \dots, N(A_k) = a_k | N_t = n] &= P\left[\sum_{i=1}^n 1_{A_l}(U_i) = a_l \text{ for } 1 \le l \le k\right] \\
&= \frac{n!}{a_1! \dots a_k!} \prod_{l=1}^k \left[\frac{m(A_l)}{t}\right]^{a_l} \\
&= \frac{n!}{t^n} \cdot \prod_{l=1}^k \frac{[m(A_l)]^{a_l}}{a_l!}
\end{aligned}$$

and therefore,

$$\begin{aligned}
&P[N(A_1)=a_1,\ldots,N(A_k)=a_k] \\
&\quad = P[N(A_1)=a_1,\ldots,N(A_k)=a_k|N_t=n]\cdot P(N_t=n) \\
&\quad = \frac{n!}{t^n}\cdot\prod_{l=1}^{k}\frac{[m(A_l)]^{a_l}}{a_l!}\cdot e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\
&\quad = \prod_{l=1}^{k}\frac{[m(A_l)]^{a_l}}{a_l!}\cdot e^{-\lambda t}\lambda^n \\
&\quad = \prod_{l=1}^{k}\frac{[m(A_l)\lambda]^{a_l}}{a_l!}e^{-\lambda a_l}
\end{aligned}$$

which shows that $\{N(A_l)\}_{l=1}^k$ are independent and that $N(A_l)\overset{d}{=}\mathrm{Poi}(\lambda m(A_l))$ for each l. ∎

Remark 11.12. If $A\in\mathcal{B}_{[0,\infty)}$ with $m(A)=\infty$, then $N(A)=\infty$ a.s. To prove this observe that $N(A)=\uparrow\lim_{n\to\infty}N(A\cap[0,n])$. Therefore for any $k\in\mathbb{N}$, we have

$$\begin{aligned}
P(N(A)\geq k) &\geq P(N(A\cap[0,n])\geq k) \\
&= 1-e^{-\lambda m(A\cap[0,n])}\sum_{0\leq l<k}\frac{(\lambda m(A\cap[0,n]))^l}{l!}\to 1 \text{ as } n\to\infty.
\end{aligned}$$

This shows that $N(A)\geq k$ a.s. for all $k\in\mathbb{N}$, i.e. $N(A)=\infty$ a.s.

Exercise 11.6 (A Generalized Poisson Process). Suppose that $(S,\mathcal{B}_S,\mu)$ is a finite measure space with $\mu(S)<\infty$. Define $\Omega=\sum_{n=0}^\infty S^n$ where $S^0=\{*\}$, were $*$ is some arbitrary point. Define $\mathcal{B}_\Omega$ to be those sets, $B=\sum_{n=0}^\infty B_n$ where $B_n\in\mathcal{B}_{S^n}:=\mathcal{B}_S^{\otimes n}$ – the product σ – algebra on S^n. Now define a probability measure, P, on $(\Omega,\mathcal{B}_\Omega)$ by

$$P(B):=e^{-\mu(S)}\sum_{n=0}^\infty\frac{1}{n!}\mu^{\otimes n}(B_n)$$

where $\mu^{\otimes 0}(\{*\})=1$ by definition. (We denote P schematically by $P:=e^{-\mu(S)}e^{\mu\otimes}$.) Finally for ever $\omega\in\Omega$, let N_ω, be the point measure on $(S,\mathcal{B}_S)$ defined by; $N_*=0$ and

$$N_\omega=\sum_{i=1}^n\delta_{s_i}\text{ if }\omega=(s_1,\ldots,s_n)\in S^n\text{ for }n\geq 1.$$

So for $A\in\mathcal{B}_S$, we have $N_*(A)=0$ and $N_\omega(A)=\sum_{i=1}^n 1_A(s_i)$. Show;

1. For each $A\in\mathcal{B}_S$, $\omega\to N_\omega(A)$ is a Poisson random variable with intensity $\mu(A)$, i.e. $N(A)=\mathrm{Poi}(\mu(A))$.
2. If $\{A_k\}_{k=1}^m\subset\mathcal{B}_S$ are disjoint sets, the $\{\omega\to N_\omega(A_k)\}_{k=1}^m$ are independent random variables.

The random measure, $\omega\to N_\omega$ is called a **Poisson process on** $(S,\mathcal{B}_S)$ **with intensity measure** μ.

11.4 Poission Process Extras*

(This subsection still needs work!) In Definition 11.7 we really gave a construction of a Poisson process as defined in Definition 11.13. The goal of this section is to show that the Poisson process, $\{N_t\}_{t\geq 0}$, as defined in Definition 11.13 is uniquely determined and is essentially equivalent to what we have already done above.

Definition 11.13 (Poisson Process II). *Let $(\Omega,\mathcal{B},P)$ be a probability space and $N_t:\Omega\to\mathbb{N}_0$ be a random variable for each $t\geq 0$. We say that $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ if; 1) $N_0=0$, 2) $N_t-N_s\overset{d}{=}\mathrm{Poi}(\lambda(t-s))$ for all $0\leq s<t<\infty$, 3) $\{N_t\}_{t\geq 0}$ has independent increments, and 4) $t\to N_t(\omega)$ is right continuous and non-decreasing for all $\omega\in\Omega$.*

Let $N_\infty(\omega):=\uparrow\lim_{t\uparrow\infty}N_t(\omega)$ and observe that $N_\infty=\sum_{k=0}^\infty(N_k-N_{k-1})=\infty$ a.s. by Lemma 11.2. Therefore, we may and do assume that $N_\infty(\omega)=\infty$ for all $\omega\in\Omega$.

Lemma 11.14. *There is zero probability that $\{N_t\}_{t\geq 0}$ makes a jump greater than or equal to 2.*

Proof. Suppose that $T\in(0,\infty)$ is fixed and $\omega\in\Omega$ is sample point where $t\to N_t(\omega)$ makes a jump of 2 or more for $t\in[0,T]$. Then for all $n\in\mathbb{N}$ we must have $\omega\in\cup_{k=1}^n\left\{N_{\frac{k}{n}T}-N_{\frac{k-1}{n}T}\geq 2\right\}$. Therefore,

$$\begin{aligned}
&P^*(\{\omega:[0,T]\ni t\to N_t(\omega)\text{ has jump }\geq 2\}) \\
&\quad\leq\sum_{k=1}^n P\left(N_{\frac{k}{n}T}-N_{\frac{k-1}{n}T}\geq 2\right)=\sum_{k=1}^n O\left(T^2/n^2\right)=O(1/n)\to 0
\end{aligned}$$

as $n\to\infty$. I am leaving open the possibility that the set of ω where a jump size 2 or larger is not measurable. ∎

Theorem 11.15. *Suppose that $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ as in Definition 11.13,*

$$W_n := \inf\{t : N_t = n\} \text{ for all } n \in \mathbb{N}_0$$

be the first time N_t reaches n. (The $\{W_n\}_{n=0}^{\infty}$ are well defined off a set of measure zero and $W_n < W_{n+1}$ for all n by the right continuity of $\{N_t\}_{t\geq 0}$.) Then the $\{T_n := W_n - W_{n-1}\}_{n=1}^{\infty}$ are i.i.d. $E(\lambda)$ – random variables. Thus the two descriptions of a Poisson process given in Definitions 11.7 and 11.13 are equivalent.

Proof. Suppose that $J_i = (a_i, b_i]$ with $b_i \leq a_{i+1} < \infty$ for all i. We will begin by showing

$$P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) = \lambda^n \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} e^{-\lambda w_n} dw_n \tag{11.7}$$

$$= \lambda^n \int_{J_1 \times J_2 \times \dots \times J_n} e^{-\lambda w_n} dw_1 \dots dw_n. \tag{11.8}$$

To show this let $K_i := (b_{i-1}, a_i]$ where $b_0 = 0$. Then

$$\cap_{i=1}^n \{W_i \in J_i\} = \cap_{i=1}^n \{N(K_i) = 0\} \cap \cap_{i=1}^{n-1} \{N(J_i) = 0\} \cap \{N(J_n) \geq 2\}$$

and therefore,

$$\begin{aligned} P\left(\cap_{i=1}^n \{W_i \in J_i\}\right) &= \prod_{i=1}^{n} e^{-\lambda m(K_i)} \cdot \prod_{i=1}^{n-1} e^{-\lambda m(J_i)} \lambda m(J_i) \cdot \left(1 - e^{-\lambda m(J_n)}\right) \\ &= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \left[e^{-\lambda a_n} - e^{-\lambda b_n}\right] \\ &= \lambda^{n-1} \prod_{i=1}^{n-1} m(J_i) \cdot \int_{J_n} \lambda e^{-\lambda w_n} dw_n. \end{aligned}$$

We may now apply a $\pi - \lambda$ – argument, using $\sigma(\{J_1 \times \dots \times J_n\}) = \mathcal{B}_{\Delta_n}$, to show

$$\mathbb{E}\left[g(W_1, \dots, W_n)\right] = \int_{\Delta_n} g(w_1, \dots, w_n) \lambda^n e^{-\lambda w_n} dw_1 \dots dw_n$$

holds for all bounded $\mathcal{B}_{\Delta_n}/\mathcal{B}_{\mathbb{R}}$ measurable functions, $g : \Delta_n \to \mathbb{R}$. Undoing the change of variables you made in Exercise 11.2 allows us to conclude that $\{T_n\}_{n=1}^{\infty}$ are i.i.d. $E(\lambda)$ – distributed random variables. ∎

L^p – spaces

Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and for $0 < p < \infty$ and a measurable function $f : \Omega \to \mathbb{C}$ let

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} \tag{12.1}$$

and when $p = \infty$, let

$$\|f\|_\infty = \inf \{a \ge 0 : \mu(|f| > a) = 0\} \tag{12.2}$$

For $0 < p \le \infty$, let

$$L^p(\Omega, \mathcal{B}, \mu) = \{f : \Omega \to \mathbb{C} : f \text{ is measurable and } \|f\|_p < \infty\} / \sim$$

where $f \sim g$ iff $f = g$ a.e. Notice that $\|f - g\|_p = 0$ iff $f \sim g$ and if $f \sim g$ then $\|f\|_p = \|g\|_p$. In general we will (by abuse of notation) use f to denote both the function f and the equivalence class containing f.

Remark 12.1. Suppose that $\|f\|_\infty \le M$, then for all $a > M$, $\mu(|f| > a) = 0$ and therefore $\mu(|f| > M) = \lim_{n\to\infty} \mu(|f| > M + 1/n) = 0$, i.e. $|f(\omega)| \le M$ for μ - a.e. ω. Conversely, if $|f| \le M$ a.e. and $a > M$ then $\mu(|f| > a) = 0$ and hence $\|f\|_\infty \le M$. This leads to the identity:

$$\|f\|_\infty = \inf \{a \ge 0 : |f(\omega)| \le a \text{ for } \mu - \text{a.e. } \omega\}.$$

12.1 Modes of Convergence

Let $\{f_n\}_{n=1}^\infty \cup \{f\}$ be a collection of complex valued measurable functions on Ω. We have the following notions of convergence and Cauchy sequences.

Definition 12.2.
1. *$f_n \to f$ a.e. if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $\lim_{n\to\infty} 1_{E^c} f_n = 1_{E^c} f$.*
2. *$f_n \to f$ in μ – measure if $\lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0$ for all $\varepsilon > 0$. We will abbreviate this by saying $f_n \to f$ in L^0 or by $f_n \xrightarrow{\mu} f$.*
3. *$f_n \to f$ in L^p iff $f \in L^p$ and $f_n \in L^p$ for all n, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.*

Definition 12.3.
1. *$\{f_n\}$ is a.e. Cauchy if there is a set $E \in \mathcal{B}$ such that $\mu(E) = 0$ and $\{1_{E^c} f_n\}$ is a pointwise Cauchy sequences.*
2. *$\{f_n\}$ is Cauchy in μ – measure (or L^0 – Cauchy) if $\lim_{m,n\to\infty} \mu(|f_n - f_m| > \varepsilon) = 0$ for all $\varepsilon > 0$.*
3. *$\{f_n\}$ is Cauchy in L^p if $\lim_{m,n\to\infty} \|f_n - f_m\|_p = 0$.*

When μ is a probability measure, we describe, $f_n \xrightarrow{\mu} f$ as f_n **converging to f in probability.** If a sequence $\{f_n\}_{n=1}^\infty$ is L^p – convergent, then it is L^p – Cauchy. For example, when $p \in [1, \infty]$ and $f_n \to f$ in L^p, we have

$$\|f_n - f_m\|_p \le \|f_n - f\|_p + \|f - f_m\|_p \to 0 \text{ as } m, n \to \infty.$$

The case where $p = 0$ will be handled in Theorem 12.7 below.

Lemma 12.4 (L^p – convergence implies convergence in probability). *Let $p \in [1, \infty)$. If $\{f_n\} \subset L^p$ is L^p – convergent (Cauchy) then $\{f_n\}$ is also convergent (Cauchy) in measure.*

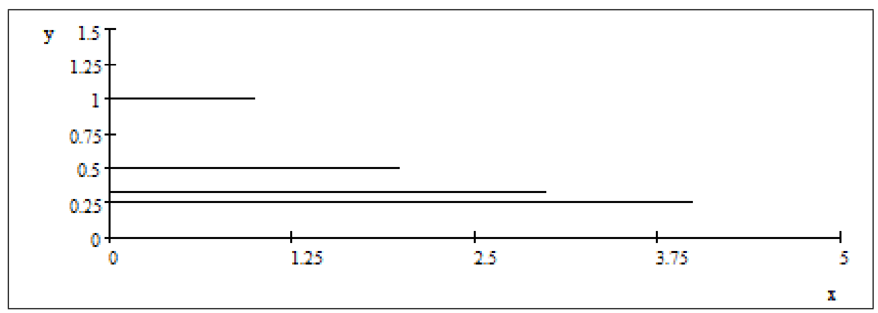
Proof. By Chebyshev's inequality (7.2),

$$\mu(|f| \ge \varepsilon) = \mu(|f|^p \ge \varepsilon^p) \le \frac{1}{\varepsilon^p} \int_\Omega |f|^p \, d\mu = \frac{1}{\varepsilon^p} \|f\|_p^p$$

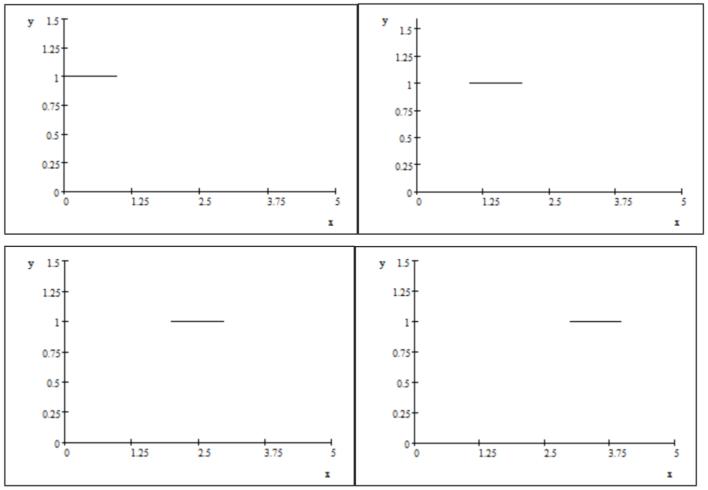
and therefore if $\{f_n\}$ is L^p – Cauchy, then

$$\mu(|f_n - f_m| \ge \varepsilon) \le \frac{1}{\varepsilon^p} \|f_n - f_m\|_p^p \to 0 \text{ as } m, n \to \infty$$

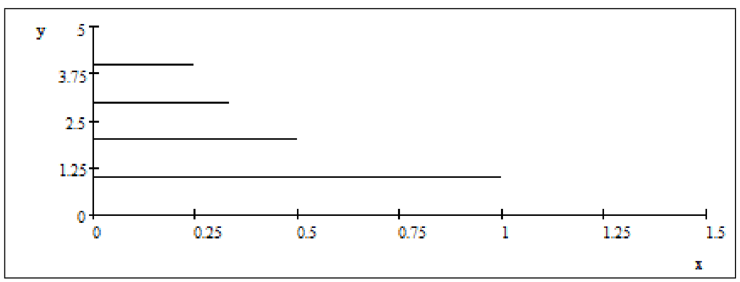
showing $\{f_n\}$ is L^0 – Cauchy. A similar argument holds for the L^p – convergent case. ∎



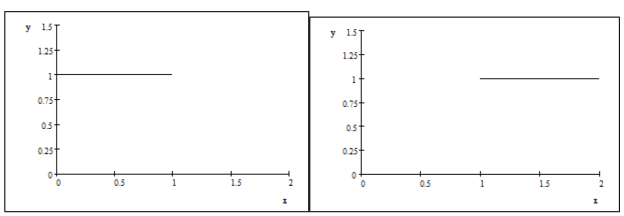
Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \nrightarrow 0$ in L^1, $f_n \xrightarrow{m} 0$.

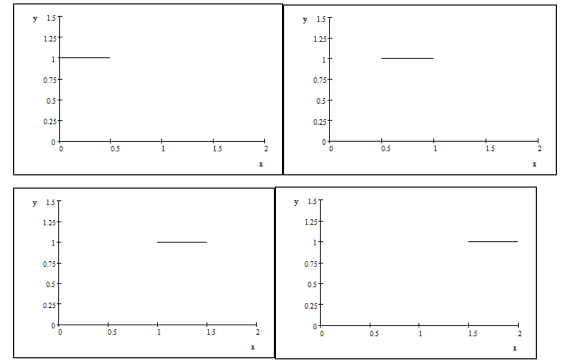


Above is a sequence of functions where $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in L^1. or in measure.



Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \overset{m}{\to} 0$ but $f_n \nrightarrow 0$ in L^1.





Above is a sequence of functions where $f_n \to 0$ in L^1, $f_n \nrightarrow 0$ a.e., and $f_n \overset{m}{\to} 0$.

Theorem 12.5 (Egoroff: *a.s.* $\implies$ **convergence in probability).** *Suppose $\mu(\Omega) = 1$ and $f_n \to f$ a.s. Then for all $\varepsilon > 0$ there exists $E = E_\varepsilon \in \mathcal{B}$ such that $\mu(E) < \varepsilon$ and $f_n \to f$ uniformly on E^c. In particular $f_n \xrightarrow{\mu} f$ as $n \to \infty$.*

Proof. Let $f_n \to f$ a.e. Then for all $\varepsilon > 0$,

$$\begin{aligned} 0 &= \mu(\{|f_n - f| > \varepsilon \text{ i.o. } n\}) \\ &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} \{|f_n - f| > \varepsilon\}\right) \qquad (12.3) \\ &\geq \limsup_{N\to\infty} \mu\left(\{|f_N - f| > \varepsilon\}\right) \end{aligned}$$

from which it follows that $f_n \xrightarrow{\mu} f$ as $n \to \infty$. To get the uniform convergence off a small exceptional set, the equality in Eq. (12.3) allows us to choose an increasing sequence $\{N_k\}_{k=1}^\infty$, such that, if

$$E_k := \bigcup_{n\geq N_k} \left\{|f_n - f| > \frac{1}{k}\right\}, \text{ then } \mu(E_k) < \varepsilon 2^{-k}.$$

The set, $E := \cup_{k=1}^\infty E_k$, then satisfies the estimate, $\mu(E) < \sum_k \varepsilon 2^{-k} = \varepsilon$. Moreover, for $\omega \notin E$, we have $|f_n(\omega) - f(\omega)| \leq \frac{1}{k}$ for all $n \geq N_k$ and all k. That is $f_n \to f$ uniformly on E^c. ■

Lemma 12.6. *Suppose $a_n \in \mathbb{C}$ and $|a_{n+1} - a_n| \leq \varepsilon_n$ and $\sum_{n=1}^\infty \varepsilon_n < \infty$. Then $\lim_{n\to\infty} a_n = a \in \mathbb{C}$ exists and $|a - a_n| \leq \delta_n := \sum_{k=n}^\infty \varepsilon_k$.*

Proof. Let $m > n$ then

$$|a_m - a_n| = \left|\sum_{k=n}^{m-1} (a_{k+1} - a_k)\right| \le \sum_{k=n}^{m-1} |a_{k+1} - a_k| \le \sum_{k=n}^{\infty} \varepsilon_k := \delta_n . \qquad (12.4)$$

So $|a_m - a_n| \le \delta_{\min(m,n)} \to 0$ as $, m, n \to \infty$, i.e. $\{a_n\}$ is Cauchy. Let $m \to \infty$ in (12.4) to find $|a - a_n| \le \delta_n$. ∎

Theorem 12.7. *Let $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions on Ω.*

1. *If f and g are measurable functions and $f_n \xrightarrow{\mu} f$ and $f_n \xrightarrow{\mu} g$ then $f = g$ a.e.*
2. *If $f_n \xrightarrow{\mu} f$ then $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure.*
3. *If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure, there exists a measurable function, f, and a subsequence $g_j = f_{n_j}$ of $\{f_n\}$ such that $\lim_{j\to\infty} g_j := f$ exists a.e.*
4. *If $\{f_n\}_{n=1}^{\infty}$ is Cauchy in measure and f is as in item 3. then $f_n \xrightarrow{\mu} f$.*
5. *Let us now further assume that $\mu(\Omega) < \infty$. In this case, a sequence of functions, $\{f_n\}_{n=1}^{\infty}$ converges to f in probability iff every subsequence, $\{f_n'\}_{n=1}^{\infty}$ of $\{f_n\}_{n=1}^{\infty}$ has a further subsequence, $\{f_n''\}_{n=1}^{\infty}$, which is almost surely convergent to f.*

Proof.

1. Suppose that f and g are measurable functions such that $f_n \xrightarrow{\mu} g$ and $f_n \xrightarrow{\mu} f$ as $n \to \infty$ and $\varepsilon > 0$ is given. Since

$$\begin{aligned}\{|f - g| > \varepsilon\} &= \{|f - f_n + f_n - g| > \varepsilon\} \subset \{|f - f_n| + |f_n - g| > \varepsilon\} \\ &\subset \{|f - f_n| > \varepsilon/2\} \cup \{|g - f_n| > \varepsilon/2\},\end{aligned}$$

$$\mu(|f - g| > \varepsilon) \le \mu(|f - f_n| > \varepsilon/2) + \mu(|g - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty.$$

Hence

$$\mu(|f - g| > 0) = \mu\left(\cup_{n=1}^{\infty}\left\{|f - g| > \frac{1}{n}\right\}\right) \le \sum_{n=1}^{\infty} \mu\left(|f - g| > \frac{1}{n}\right) = 0,$$

i.e. $f = g$ a.e.

2. Suppose $f_n \xrightarrow{\mu} f$, $\varepsilon > 0$ and $m, n \in \mathbb{N}$ and $\omega \in \Omega$ are such that $|f_n(\omega) - f_m(\omega)| > \varepsilon$. Then

$$\varepsilon < |f_n(\omega) - f_m(\omega)| \le |f_n(\omega) - f(\omega)| + |f(\omega) - f_m(\omega)|$$

from which it follows that either $|f_n(\omega) - f(\omega)| > \varepsilon/2$ or $|f(\omega) - f_m(\omega)| > \varepsilon/2$. Therefore we have shown,

$$\{|f_n - f_m| > \varepsilon\} \subset \{|f_n - f| > \varepsilon/2\} \cup \{|f_m - f| > \varepsilon/2\}$$

and hence

$$\mu(|f_n - f_m| > \varepsilon) \le \mu(|f_n - f| > \varepsilon/2) + \mu(|f_m - f| > \varepsilon/2) \to 0 \text{ as } m, n \to \infty.$$

3. Suppose $\{f_n\}$ is $L^0(\mu)$ – Cauchy and let $\varepsilon_n > 0$ such that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum_{k=n}^{\infty} \varepsilon_k$. Choose $g_j = f_{n_j}$ where $\{n_j\}$ is a subsequence of $\mathbb{N}$ such that

$$\mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \le \varepsilon_j.$$

Let $F_N := \cup_{j \ge N} \{|g_{j+1} - g_j| > \varepsilon_j\}$ and

$$E := \cap_{N=1}^{\infty} F_N = \{|g_{j+1} - g_j| > \varepsilon_j \text{ i.o.}\}$$

and observe that $\mu(F_N) \le \delta_N < \infty$. Since

$$\sum_{j=1}^{\infty} \mu(\{|g_{j+1} - g_j| > \varepsilon_j\}) \le \sum_{j=1}^{\infty} \varepsilon_j < \infty,$$

it follows from the first Borel-Cantelli lemma that

$$0 = \mu(E) = \lim_{N\to\infty} \mu(F_N).$$

For $\omega \notin E$, $|g_{j+1}(\omega) - g_j(\omega)| \le \varepsilon_j$ for a.a. j and so by Lemma 12.6, $f(\omega) := \lim_{j\to\infty} g_j(\omega)$ exists. For $\omega \in E$ we may define $f(\omega) \equiv 0$.

4. Next we will show $g_N \xrightarrow{\mu} f$ as $N \to \infty$ where f and g_N are as above. If

$$\omega \in F_N^c = \cap_{j \ge N} \{|g_{j+1} - g_j| \le \varepsilon_j\},$$

then

$$|g_{j+1}(\omega) - g_j(\omega)| \le \varepsilon_j \text{ for all } j \ge N.$$

Another application of Lemma 12.6 shows $|f(\omega) - g_j(\omega)| \le \delta_j$ for all $j \ge N$, i.e.

$$F_N^c \subset \cap_{j \ge N} \{\omega \in \Omega : |f(\omega) - g_j(\omega)| \le \delta_j\}.$$

Taking complements of this equation shows

$$\{|f - g_N| > \delta_N\} \subset \cup_{j \ge N} \{|f - g_j| > \delta_j\} \subset F_N.$$

and therefore,

$$\mu(|f-g_N|>\delta_N)\leq\mu(F_N)\leq\delta_N\to 0 \text{ as } N\to\infty$$

and in particular, $g_N \overset{\mu}{\to} f$ as $N\to\infty$.
With this in hand, it is straightforward to show $f_n \overset{\mu}{\to} f$. Indeed, since

$$\begin{aligned}\{|f_n-f|>\varepsilon\} &= \{|f-g_j+g_j-f_n|>\varepsilon\}\\ &\subset \{|f-g_j|+|g_j-f_n|>\varepsilon\}\\ &\subset \{|f-g_j|>\varepsilon/2\}\cup\{|g_j-f_n|>\varepsilon/2\},\end{aligned}$$

we have

$$\mu(\{|f_n-f|>\varepsilon\})\leq\mu(\{|f-g_j|>\varepsilon/2\})+\mu(|g_j-f_n|>\varepsilon/2).$$

Therefore, letting $j\to\infty$ in this inequality gives,

$$\mu(\{|f_n-f|>\varepsilon\})\leq\limsup_{j\to\infty}\mu(|g_j-f_n|>\varepsilon/2)\to 0 \text{ as } n\to\infty$$

because $\{f_n\}_{n=1}^\infty$ was Cauchy in measure.

5. If $\{f_n\}_{n=1}^\infty$ is convergent and hence Cauchy in probability then any subsequence, $\{f_n'\}_{n=1}^\infty$ is also Cauchy in probability. Hence by item 3. there is a further subsequence, $\{f_n''\}_{n=1}^\infty$ of $\{f_n'\}_{n=1}^\infty$ which is convergent almost surely. Conversely if $\{f_n\}_{n=1}^\infty$ does not converge to f in probability, then there exists an $\varepsilon>0$ and a subsequence, $\{n_k\}$ such that $\inf_k \mu(|f-f_{n_k}|\geq\varepsilon)>0$. Any subsequence of $\{f_{n_k}\}$ would have the same property and hence can not be almost surely convergent because of Theorem 12.5.

■

Corollary 12.8 (Dominated Convergence Theorem). *Let $(\Omega,\mathcal{B},\mu)$ be a measure space. Suppose $\{f_n\}$, $\{g_n\}$, and g are in L^1 and $f\in L^0$ are functions such that*

$$|f_n|\leq g_n \text{ a.e., } f_n\overset{\mu}{\longrightarrow}f,\ g_n\overset{\mu}{\longrightarrow}g, \text{ and } \int g_n\to\int g \text{ as } n\to\infty.$$

Then $f\in L^1$ and $\lim_{n\to\infty}\|f-f_n\|_1=0$, i.e. $f_n\to f$ in L^1. In particular $\lim_{n\to\infty}\int f_n=\int f$.

Proof. First notice that $|f|\leq g$ a.e. and hence $f\in L^1$ since $g\in L^1$. To see that $|f|\leq g$, use Theorem 12.7 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f|=\lim_{k\to\infty}|f_{n_k}|\leq\lim_{k\to\infty}g_{n_k}=g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty}\|f-f_n\|_1\neq 0$ there exists $\varepsilon>0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int|f-f_{n_k}|\geq\varepsilon \text{ for all } k. \tag{12.5}$$

Using Theorem 12.7 again, we may assume (by passing to a further subsequences if necessary) that $f_{n_k}\to f$ and $g_{n_k}\to g$ almost everywhere. Noting, $|f-f_{n_k}|\leq g+g_{n_k}\to 2g$ and $\int(g+g_{n_k})\to\int 2g$, an application of the dominated convergence Theorem 7.27 implies $\lim_{k\to\infty}\int|f-f_{n_k}|=0$ which contradicts Eq. (12.5). ■

Exercise 12.1 (Fatou's Lemma). Let $(\Omega,\mathcal{B},\mu)$ be a measure space. If $f_n\geq 0$ and $f_n\to f$ in measure, then $\int_\Omega f d\mu\leq\liminf_{n\to\infty}\int_\Omega f_n d\mu$.

Exercise 12.2. Let $(\Omega,\mathcal{B},\mu)$ be a measure space, $p\in[1,\infty)$, $\{f_n\}\subset L^p(\mu)$ and $f\in L^p(\mu)$. Then $f_n\to f$ in $L^p(\mu)$ iff $f_n\overset{\mu}{\longrightarrow}f$ and $\int|f_n|^p\to\int|f|^p$.

Solution to Exercise (12.2). By the triangle inequality, $\left|\|f\|_p-\|f_n\|_p\right|\leq\|f-f_n\|_p$ which shows $\int|f_n|^p\to\int|f|^p$ if $f_n\to f$ in L^p. Moreover Chebyschev's inequality implies $f_n\overset{\mu}{\longrightarrow}f$ if $f_n\to f$ in L^p.

For the converse, let $F_n:=|f-f_n|^p$ and $G_n:=2^{p-1}\left[|f|^p+|f_n|^p\right]$. Then $F_n\overset{\mu}{\longrightarrow}0$, $F_n\leq G_n\in L^1$, and $\int G_n\to\int G$ where $G:=2^p|f|^p\in L^1$. Therefore, by Corollary 12.8, $\int|f-f_n|^p=\int F_n\to\int 0=0$.

Exercise 12.3. Let $(\Omega,\mathcal{B},\mu)$ be a measure space, $p\in[1,\infty)$, and suppose that $0\leq f\in L^1(\mu)$, $0\leq f_n\in L^1(\mu)$ for all n, $f_n\overset{\mu}{\longrightarrow}f$, and $\int f_n d\mu\to\int f d\mu$. Then $f_n\to f$ in $L^1(\mu)$. In particular if $f,f_n\in L^p(\mu)$ and $f_n\to f$ in $L^p(\mu)$, then $|f_n|^p\to|f|^p$ in $L^1(\mu)$.

Solution to Exercise (12.3). Let $F_n:=|f-f_n|\leq f+f_n:=g_n$ and $g:=2f$. Then $u_n\overset{\mu}{\longrightarrow}0$, $g_n\overset{\mu}{\longrightarrow}g$, and $\int g_n d\mu\to\int g d\mu$. So by Corollary 12.8, $\int|f-f_n|d\mu=\int F_n d\mu\to 0$ as $n\to\infty$.

Corollary 12.9. *Suppose $(\Omega,\mathcal{B},\mu)$ is a probability space, $f_n\overset{\mu}{\longrightarrow}f$ and $g_n\overset{\mu}{\longrightarrow}g$ and $\varphi:\mathbb{R}\to\mathbb{R}$ and $\psi:\mathbb{R}^2\to\mathbb{R}$ are continuous functions. Then*

1. $\varphi(f_n)\overset{\mu}{\longrightarrow}\varphi(f)$,
2. $\psi(f_n,g_n)\overset{\mu}{\longrightarrow}\psi(f,g)$,
3. $f_n+g_n\overset{\mu}{\longrightarrow}f+g$, *and*
4. $f_n\cdot g_n\overset{\mu}{\longrightarrow}f\cdot g$.

Proof. Item 1., 3. and 4. all follow from item 2. by taking $\psi(x,y)=\varphi(x)$, $\psi(x,y)=x+y$, and $\psi(x,y)=x\cdot y$ respectively. So it suffices to prove item 2. To do this we will make repeated use of Theorem 12.7.

Given a subsequence, $\{n_k\}$, of $\mathbb{N}$ there is a subsequence, $\{n_k'\}$ of $\{n_k\}$ such that $f_{n_k'} \to f$ a.s. and yet a further subsequence $\{n_k''\}$ of $\{n_k'\}$ such that $g_{n_k''} \to g$ a.s. Hence, by the continuity of ψ, it now follows that

$$\lim_{k\to\infty} \psi\left(f_{n_k''}, g_{n_k''}\right) = \psi(f, g) \text{ a.s.}$$

which completes the proof. ■

12.2 Jensen's, Hölder's and Minikowski's Inequalities

Theorem 12.10 (Jensen's Inequality). *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a probability space, i.e. μ is a positive measure and $\mu(\Omega) = 1$. Also suppose that $f \in L^1(\mu)$, $f : \Omega \to (a, b)$, and $\varphi : (a, b) \to \mathbb{R}$ is a convex function, (i.e. $\varphi''(x) \geq 0$ on (a, b).) Then*

$$\varphi\left(\int_\Omega f d\mu\right) \leq \int_\Omega \varphi(f) d\mu$$

where if $\varphi \circ f \notin L^1(\mu)$, then $\varphi \circ f$ is integrable in the extended sense and $\int_\Omega \varphi(f) d\mu = \infty$.

Proof. Let $t = \int_\Omega f d\mu \in (a, b)$ and let $\beta \in \mathbb{R}$ ($\beta = \dot\varphi(t)$ when $\dot\varphi(t)$ exists), be such that $\varphi(s) - \varphi(t) \geq \beta(s - t)$ for all $s \in (a, b)$. (See Lemma 12.44) and Figure 12.3 when φ is C^1 and Theorem 12.47 below for the existence of such a β in the general case.) Then integrating the inequality, $\varphi(f) - \varphi(t) \geq \beta(f - t)$, implies that

$$0 \leq \int_\Omega \varphi(f) d\mu - \varphi(t) = \int_\Omega \varphi(f) d\mu - \varphi(\int_\Omega f d\mu).$$

Moreover, if $\varphi(f)$ is not integrable, then $\varphi(f) \geq \varphi(t) + \beta(f - t)$ which shows that negative part of $\varphi(f)$ is integrable. Therefore, $\int_\Omega \varphi(f) d\mu = \infty$ in this case. ■

Example 12.11. Since e^x for $x \in \mathbb{R}$, $-\ln x$ for $x > 0$, and x^p for $x \geq 0$ and $p \geq 1$ are all convex functions, we have the following inequalities

$$\exp\left(\int_\Omega f d\mu\right) \leq \int_\Omega e^f d\mu, \tag{12.6}$$

$$\int_\Omega \log(|f|) d\mu \leq \log\left(\int_\Omega |f| \, d\mu\right)$$

and for $p \geq 1$,

$$\left|\int_\Omega f d\mu\right|^p \leq \left(\int_\Omega |f| \, d\mu\right)^p \leq \int_\Omega |f|^p \, d\mu.$$

As a special case of Eq. (12.6), if $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}. \tag{12.7}$$

Indeed, we have applied Eq. (12.6) with $\Omega = \{1, 2, \ldots, n\}$, $\mu = \sum_{i=1}^n \frac{1}{p_i} \delta_i$ and $f(i) := \ln s_i^{p_i}$. As a special case of Eq. (12.7), suppose that $s, t, p, q \in (1, \infty)$ with $q = \frac{p}{p-1}$ (i.e. $\frac{1}{p} + \frac{1}{q} = 1$) then

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q. \tag{12.8}$$

(When $p = q = 1/2$, the inequality in Eq. (12.8) follows from the inequality, $0 \leq (s - t)^2$.)

As another special case of Eq. (12.7), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \ldots a_n} \leq \frac{1}{n} \sum_{i=1}^n a_i. \tag{12.9}$$

Theorem 12.12 (Hölder's inequality). *Suppose that $1 \leq p \leq \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q. \tag{12.10}$$

Assuming $p \in (1, \infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (12.10) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p \ \text{ a.e.} \tag{12.11}$$

Proof. The cases $p = 1$ and $q = \infty$ or $p = \infty$ and $q = 1$ are easy to deal with and will be left to the reader. So we now assume that $p, q \in (1, \infty)$. If $\|f\|_q = 0$ or ∞ or $\|g\|_p = 0$ or ∞, Eq. (12.10) is again easily verified. So we will now assume that $0 < \|f\|_q, \|g\|_p < \infty$. Taking $s = |f| / \|f\|_p$ and $t = |g| / \|g\|_q$ in Eq. (12.8) gives,

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{12.12}$$

with equality iff $|g / \|g\|_q| = |f|^{p-1} / \|f\|_p^{(p-1)} = |f|^{p/q} / \|f\|_p^{p/q}$, i.e. $|g|^q \|f\|_p^p = \|g\|_q^q \, |f|^p$. Integrating Eq. (12.12) implies

$$\frac{\|fg\|_1}{\|f\|_p\|g\|_q} \le \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (12.11) holds. The proof is finished since it is easily checked that equality holds in Eq. (12.10) when $|f|^p = c\,|g|^q$ of $|g|^q = c\,|f|^p$ for some constant c. ■

Example 12.13. Suppose that $a_k \in \mathbb{C}$ for $k = 1, 2, \ldots, n$ and $p \in [1, \infty)$, then

$$\left|\sum_{k=1}^n a_k\right|^p \le n^{p-1} \sum_{k=1}^n |a_k|^p . \tag{12.13}$$

Indeed, by Hölder's inequality applied using the measure space, $\{1, 2, \ldots, n\}$ equipped with counting measure, we have

$$\left|\sum_{k=1}^n a_k\right| = \left|\sum_{k=1}^n a_k \cdot 1\right| \le \left(\sum_{k=1}^n |a_k|^p\right)^{1/p} \left(\sum_{k=1}^n 1^q\right)^{1/q} = n^{1/q} \left(\sum_{k=1}^n |a_k|^p\right)^{1/p}$$

where $q = \frac{p}{p-1}$. Taking the p^{th} – power of this inequality then gives, Eq. (12.14).

Theorem 12.14 (Generalized Hölder's inequality). *Suppose that $f_i : \Omega \to \mathbb{C}$ are measurable functions for $i = 1, \ldots, n$ and $p_1, \ldots, p_n$ and r are positive numbers such that $\sum_{i=1}^n p_i^{-1} = r^{-1}$, then*

$$\left\|\prod_{i=1}^n f_i\right\|_r \le \prod_{i=1}^n \|f_i\|_{p_i} . \tag{12.14}$$

Proof. One may prove this theorem by induction based on Hölder's Theorem 12.12 above. Alternatively we may give a proof along the lines of the proof of Theorem 12.12 which is what we will do here.

Since Eq. (12.14) is easily seen to hold if $\|f_i\|_{p_i} = 0$ for some i, we will assume that $\|f_i\|_{p_i} > 0$ for all i. By assumption, $\sum_{i=1}^n \frac{r_i}{p_i} = 1$, hence we may replace s_i by s_i^r and p_i by p_i/r for each i in Eq. (12.7) to find

$$s_1^r \ldots s_n^r \le \sum_{i=1}^n \frac{(s_i^r)^{p_i/r}}{p_i/r} = r \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}.$$

Now replace s_i by $|f_i| / \|f_i\|_{p_i}$ in the previous inequality and integrate the result to find

$$\frac{1}{\prod_{i=1}^n \|f_i\|_{p_i}} \left\|\prod_{i=1}^n f_i\right\|_r^r \le r \sum_{i=1}^n \frac{1}{p_i} \frac{1}{\|f_i\|_{p_i}^{p_i}} \int_\Omega |f_i|^{p_i}\, d\mu = \sum_{i=1}^n \frac{r}{p_i} = 1.$$

■

Theorem 12.15 (Minkowski's Inequality). *If $1 \le p \le \infty$ and $f, g \in L^p$ then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p. \tag{12.15}$$

Proof. When $p = \infty$, $|f| \le \|f\|_\infty$ a.e. and $|g| \le \|g\|_\infty$ a.e. so that $|f + g| \le |f| + |g| \le \|f\|_\infty + \|g\|_\infty$ a.e. and therefore

$$\|f + g\|_\infty \le \|f\|_\infty + \|g\|_\infty .$$

When $p < \infty$,

$$|f + g|^p \le (2 \max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \le 2^p (|f|^p + |g|^p),$$

which implies[1] $f + g \in L^p$ since

$$\|f + g\|_p^p \le 2^p \left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

Furthermore, when $p = 1$ we have

$$\|f + g\|_1 = \int_\Omega |f + g| d\mu \le \int_\Omega |f|\, d\mu + \int_\Omega |g| d\mu = \|f\|_1 + \|g\|_1.$$

We now consider $p \in (1, \infty)$. We may assume $\|f + g\|_p$, $\|f\|_p$ and $\|g\|_p$ are all positive since otherwise the theorem is easily verified. Integrating

$$|f + g|^p = |f + g||f + g|^{p-1} \le (|f| + |g|)|f + g|^{p-1}$$

and then applying Holder's inequality with $q = p/(p - 1)$ gives

$$\begin{aligned} \int_\Omega |f + g|^p d\mu &\le \int_\Omega |f|\ \ |f + g|^{p-1} d\mu + \int_\Omega |g|\ \ |f + g|^{p-1} d\mu \\ &\le (\|f\|_p + \|g\|_p)\, \|\ |f + g|^{p-1}\ \|_q, \end{aligned} \tag{12.16}$$

where

$$\||f + g|^{p-1}\|_q^q = \int_\Omega (|f + g|^{p-1})^q d\mu = \int_\Omega |f + g|^p d\mu = \|f + g\|_p^p. \tag{12.17}$$

Combining Eqs. (12.16) and (12.17) implies

$$\|f + g\|_p^p \le \|f\|_p \|f + g\|_p^{p/q} + \|g\|_p \|f + g\|_p^{p/q} \tag{12.18}$$

Solving this inequality for $\|f + g\|_p$ gives Eq. (12.15). ■

[1] In light of Example 12.13, the last 2^p in the above inequality may be replaced by 2^{p-1}.

12.3 Completeness of L^p – spaces

Theorem 12.16. *Let $\|\cdot\|_\infty$ be as defined in Eq. (12.2), then $(L^\infty(\Omega,\mathcal{B},\mu),\|\cdot\|_\infty)$ is a Banach space. A sequence $\{f_n\}_{n=1}^\infty \subset L^\infty$ converges to $f \in L^\infty$ iff there exists $E \in \mathcal{B}$ such that $\mu(E)=0$ and $f_n \to f$ uniformly on E^c. Moreover, bounded simple functions are dense in L^∞.*

Proof. By Minkowski's Theorem 12.15, $\|\cdot\|_\infty$ satisfies the triangle inequality. The reader may easily check the remaining conditions that ensure $\|\cdot\|_\infty$ is a norm. Suppose that $\{f_n\}_{n=1}^\infty \subset L^\infty$ is a sequence such $f_n \to f \in L^\infty$, i.e. $\|f-f_n\|_\infty \to 0$ as $n\to\infty$. Then for all $k\in\mathbb{N}$, there exists $N_k<\infty$ such that

$$\mu\left(|f-f_n|>k^{-1}\right)=0 \text{ for all } n\ge N_k.$$

Let

$$E=\cup_{k=1}^\infty \cup_{n\ge N_k}\left\{|f-f_n|>k^{-1}\right\}.$$

Then $\mu(E)=0$ and for $x\in E^c$, $|f(x)-f_n(x)|\le k^{-1}$ for all $n\ge N_k$. This shows that $f_n\to f$ uniformly on E^c. Conversely, if there exists $E\in\mathcal{B}$ such that $\mu(E)=0$ and $f_n\to f$ uniformly on E^c, then for any $\varepsilon>0$,

$$\mu\left(|f-f_n|\ge\varepsilon\right)=\mu\left(\{|f-f_n|\ge\varepsilon\}\cap E^c\right)=0$$

for all n sufficiently large. That is to say $\limsup_{j\to\infty}\|f-f_n\|_\infty\le\varepsilon$ for all $\varepsilon>0$. The density of simple functions follows from the approximation Theorem 6.39. So the last item to prove is the completeness of L^∞.

Suppose $\varepsilon_{m,n}:=\|f_m-f_n\|_\infty\to 0$ as $m,n\to\infty$. Let $E_{m,n}=\{|f_n-f_m|>\varepsilon_{m,n}\}$ and $E:=\cup E_{m,n}$, then $\mu(E)=0$ and

$$\sup_{x\in E^c}|f_m(x)-f_n(x)|\le\varepsilon_{m,n}\to 0 \text{ as } m,n\to\infty.$$

Therefore, $f:=\lim_{n\to\infty}f_n$ exists on E^c and the limit is uniform on E^c. Letting $f=\lim_{n\to\infty}1_{E^c}f_n$, it then follows that $\lim_{n\to\infty}\|f_n-f\|_\infty=0$. ∎

Theorem 12.17 (Completeness of $L^p(\mu)$). *For $1\le p\le\infty$, $L^p(\mu)$ equipped with the L^p – norm, $\|\cdot\|_p$ (see Eq. (12.1)), is a Banach space.*

Proof. By Minkowski's Theorem 12.15, $\|\cdot\|_p$ satisfies the triangle inequality. As above the reader may easily check the remaining conditions that ensure $\|\cdot\|_p$ is a norm. So we are left to prove the completeness of $L^p(\mu)$ for $1\le p<\infty$, the case $p=\infty$ being done in Theorem 12.16.

Let $\{f_n\}_{n=1}^\infty\subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 12.4), $\{f_n\}$ is L^0-Cauchy (i.e. Cauchy in measure) and by Theorem 12.7 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j\to f$ a.e. By Fatou's Lemma,

$$\|g_j-f\|_p^p=\int\lim_{k\to\infty}\inf|g_j-g_k|^p d\mu\le\lim_{k\to\infty}\inf\int|g_j-g_k|^p d\mu$$
$$=\lim_{k\to\infty}\inf\|g_j-g_k\|_p^p\to 0 \text{ as } j\to\infty.$$

In particular, $\|f\|_p\le\|g_j-f\|_p+\|g_j\|_p<\infty$ so the $f\in L^p$ and $g_j\xrightarrow{L^p}f$. The proof is finished because,

$$\|f_n-f\|_p\le\|f_n-g_j\|_p+\|g_j-f\|_p\to 0 \text{ as } j,n\to\infty.$$

∎

See Proposition **??** for an important example of the use of this theorem. To end this section we are going to record a few results we will need later regarding subspace of $L^p(\mu)$ which are induced by sub – σ – algebras, $\mathcal{B}_0\subset\mathcal{B}$.

Lemma 12.18. *Let $(\Omega,\mathcal{B},\mu)$ be a measure space and $\mathcal{B}_0$ be a sub – σ – algebra of $\mathcal{B}$. Then for $1\le p<\infty$, the map $i:L^p(\Omega,\mathcal{B}_0,\mu)\to L^p(\Omega,\mathcal{B},\mu)$ defined by $i([f]_0)=[f]$ is a well defined linear isometry. Here we are writing,*

$$[f]_0=\{g\in \mathrm{L}^p(\Omega,\mathcal{B}_0,\mu):g=f \text{ a.e.}\} \text{ and}$$
$$[f]=\{g\in \mathrm{L}^p(\Omega,\mathcal{B},\mu):g=f \text{ a.e.}\}.$$

Moreover the image of i, $i(L^p(\Omega,\mathcal{B}_0,\mu))$, is a closed subspace of $L^p(\Omega,\mathcal{B},\mu)$.

Proof. This is proof is routine and most of it will be left to the reader. Let us just check that $i(L^p(\Omega,\mathcal{B}_0,\mu))$, is a closed subspace of $L^p(\Omega,\mathcal{B},\mu)$. To this end, suppose that $i([f_n]_0)=[f_n]$ is a convergent sequence in $\mathrm{L}^p(\Omega,\mathcal{B},\mu)$. Because, i, is an isometry it follows that $\{[f_n]_0\}_{n=1}^\infty$ is a Cauchy and hence convergent sequence in $L^p(\Omega,\mathcal{B}_0,\mu)$. Letting $f\in\mathrm{L}^p(\Omega,\mathcal{B}_0,\mu)$ such that $\|f-f_n\|_{L^p(\mu)}\to 0$, we will have, since i is isometric, that $[f_n]\to[f]=i([f]_0)\in i(L^p(\Omega,\mathcal{B}_0,\mu))$ as desired. ∎

Exercise 12.4. Let $(\Omega,\mathcal{B},\mu)$ be a measure space and $\mathcal{B}_0$ be a sub – σ – algebra of $\mathcal{B}$. Further suppose that to every $B\in\mathcal{B}$ there exists $A\in\mathcal{B}_0$ such that $\mu(B\Delta A)=0$. Show for all $1\le p<\infty$ that $i(L^p(\Omega,\mathcal{B}_0,\mu))=L^p(\Omega,\mathcal{B},\mu)$, i.e. to each $f\in\mathrm{L}^p(\Omega,\mathcal{B},\mu)$ there exists a $g\in\mathrm{L}^p(\Omega,\mathcal{B}_0,\mu)$ such that $f=g$ a.e. **Hints:** 1. verify the last assertion for simple functions in $\mathrm{L}^p(\Omega,\mathcal{B}_0,\mu)$. 2. then make use of Theorem 6.39 and Exercise 6.4.

Exercise 12.5. Suppose that $1\le p<\infty$, $(\Omega,\mathcal{B},\mu)$ is a σ – finite measure space and $\mathcal{B}_0$ is a sub – σ – algebra of $\mathcal{B}$. Show that $i(L^p(\Omega,\mathcal{B}_0,\mu))=L^p(\Omega,\mathcal{B},\mu)$ implies; to every $B\in\mathcal{B}$ there exists $A\in\mathcal{B}_0$ such that $\mu(B\Delta A)=0$.

Solution to Exercise (12.5). Let $B\in\mathcal{B}$ with $\mu(B)<\infty$. Then $1_B\in\mathrm{L}^p(\Omega,\mathcal{B},\mu)$ and hence by assumption there exists $g\in\mathrm{L}^p(\Omega,\mathcal{B}_0,\mu)$ such that

$g = 1_B$ a.e. Let $A := \{g = 1\} \in \mathcal{B}_0$ and observe that $A\Delta B \subset \{g \neq 1_B\}$. Therefore $\mu(A\Delta B) = \mu(g \neq 1_B) = 0$. For general the case we use the fact that $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space to conclude that each $B \in \mathcal{B}$ may be written as a disjoint union, $B = \sum_{n=1}^\infty B_n$, with $B_n \in \mathcal{B}$ and $\mu(B_n) < \infty$. By what we have just proved we may find $A_n \in \mathcal{B}_0$ such that $\mu(B_n \Delta A_n) = 0$. I now claim that $A := \cup_{n=1}^\infty A_n \in \mathcal{B}_0$ satisfies $\mu(A\Delta B) = 0$. Indeed, notice that

$$A \setminus B = \cup_{n=1}^\infty A_n \setminus B \subset \cup_{n=1}^\infty A_n \setminus B_n,$$

similarly $B \setminus A \subset \cup_{n=1}^\infty B_n \setminus A_n$, and therefore $A\Delta B \subset \cup_{n=1}^\infty A_n \Delta B_n$. Therefore by sub-additivity of μ, $\mu(A\Delta B) \leq \sum_{n=1}^\infty \mu(A_n \Delta B_n) = 0$.

Convention: From now on we will drop the cumbersome notation and simply identify $[f]$ with f and $L^p(\Omega, \mathcal{B}_0, \mu)$ with its image, $i(L^p(\Omega, \mathcal{B}_0, \mu))$, in $L^p(\Omega, \mathcal{B}, \mu)$.

12.4 Density Results

Theorem 12.19 (Density Theorem). *Let $p \in [1, \infty)$, $(\Omega, \mathcal{B}, \mu)$ be a measure space and $\mathbb{M}$ be an algebra of bounded $\mathbb{R}$ – valued measurable functions such that*

1. *$\mathbb{M} \subset L^p(\mu, \mathbb{R})$ and $\sigma(\mathbb{M}) = \mathcal{B}$.*
2. *There exists $\psi_k \in \mathbb{M}$ such that $\psi_k \to 1$ boundedly.*

Then to every function $f \in L^p(\mu, \mathbb{R})$, there exist $\varphi_n \in \mathbb{M}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(\mu)} = 0$, i.e. $\mathbb{M}$ is dense in $L^p(\mu, \mathbb{R})$.

Proof. Fix $k \in \mathbb{N}$ for the moment and let $\mathbb{H}$ denote those bounded $\mathcal{B}$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{M}$ such that $\lim_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{M} \subset \mathbb{H}$ and $\mathbb{H}$ is closed under bounded convergence. To verify the latter assertion, suppose $f_n \in \mathbb{H}$ and $f_n \to f$ boundedly. Then, by the dominated convergence theorem, $\lim_{n\to\infty} \|\psi_k(f - f_n)\|_{L^p(\mu)} = 0$.[2] (Take the dominating function to be $g = [2C|\psi_k|]^p$ where C is a constant bounding all of the $\{|f_n|\}_{n=1}^\infty$.) We may now choose $\varphi_n \in \mathbb{M}$ such that $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \leq \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} \leq \limsup_{n\to\infty} \|\psi_k(f - f_n)\|_{L^p(\mu)} + \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \qquad (12.19)$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 8.16, now shows $\mathbb{H}$ contains all bounded measurable functions on Ω. Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f|\leq k\}} f - f\right\|_{L^p(\mu)} = 0$. (Take the dominating function to be $g = [2C|f|]^p$ where C is a bound on all of the $|\psi_k|$.) Using this and what we have just proved, there exists $\varphi_k \in \mathbb{M}$ such that

$$\left\|\psi_k 1_{\{|f|\leq k\}} f - \varphi_k\right\|_{L^p(\mu)} \leq \frac{1}{k}.$$

The same line of reasoning used in Eq. (12.19) now implies $\lim_{k\to\infty} \|f - \varphi_k\|_{L^p(\mu)} = 0$. ■

Example 12.20. Let μ be a measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ such that $\mu([-M, M]) < \infty$ for all $M < \infty$. Then, $C_c(\mathbb{R}, \mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ with compact support) is dense in $L^p(\mu)$ for all $1 \leq p < \infty$. To see this, apply Theorem 12.19 with $\mathbb{M} = C_c(\mathbb{R}, \mathbb{R})$ and $\psi_k := 1_{[-k,k]}$.

Theorem 12.21. *Suppose $p \in [1, \infty)$, $\mathcal{A} \subset \mathcal{B} \subset 2^\Omega$ is an algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Let $\mathbb{S}(\mathcal{A}, \mu)$ denote the measurable simple functions, $\varphi : \Omega \to \mathbb{R}$ such $\{\varphi = y\} \in \mathcal{A}$ for all $y \in \mathbb{R}$ and $\mu(\{\varphi \neq 0\}) < \infty$. Then $\mathbb{S}(\mathcal{A}, \mu)$ is dense subspace of $L^p(\mu)$.*

Proof. Let $\mathbb{M} := \mathbb{S}(\mathcal{A}, \mu)$. By assumption there exists $\Omega_k \in \mathcal{A}$ such that $\mu(\Omega_k) < \infty$ and $\Omega_k \uparrow \Omega$ as $k \to \infty$. If $A \in \mathcal{A}$, then $\Omega_k \cap A \in \mathcal{A}$ and $\mu(\Omega_k \cap A) < \infty$ so that $1_{\Omega_k \cap A} \in \mathbb{M}$. Therefore $1_A = \lim_{k\to\infty} 1_{\Omega_k \cap A}$ is $\sigma(\mathbb{M})$ – measurable for every $A \in \mathcal{A}$. So we have shown that $\mathcal{A} \subset \sigma(\mathbb{M}) \subset \mathcal{B}$ and therefore $\mathcal{B} = \sigma(\mathcal{A}) \subset \sigma(\mathbb{M}) \subset \mathcal{B}$, i.e. $\sigma(\mathbb{M}) = \mathcal{B}$. The theorem now follows from Theorem 12.19 after observing $\psi_k := 1_{\Omega_k} \in \mathbb{M}$ and $\psi_k \to 1$ boundedly. ■

Theorem 12.22 (Separability of L^p – Spaces). *Suppose, $p \in [1, \infty)$, $\mathcal{A} \subset \mathcal{B}$ is a countable algebra such that $\sigma(\mathcal{A}) = \mathcal{B}$ and μ is σ – finite on $\mathcal{A}$. Then $L^p(\mu)$ is separable and*

$$\mathbb{D} = \{\sum a_j 1_{A_j} : a_j \in \mathbb{Q} + i\mathbb{Q},\ A_j \in \mathcal{A} \text{ with } \mu(A_j) < \infty\}$$

is a countable dense subset.

Proof. It is left to reader to check $\mathbb{D}$ is dense in $\mathbb{S}(\mathcal{A}, \mu)$ relative to the $L^p(\mu)$ – norm. Once this is done, the proof is then complete since $\mathbb{S}(\mathcal{A}, \mu)$ is a dense subspace of $L^p(\mu)$ by Theorem 12.21. ■

[2] It is at this point that the proof would break down if $p = \infty$.

12.5 Relationships between different L^p – spaces

The $L^p(\mu)$ – norm controls two types of behaviors of f, namely the "behavior at infinity" and the behavior of "local singularities." So in particular, if f blows up at a point $x_0 \in \Omega$, then locally near x_0 it is harder for f to be in $L^p(\mu)$ as p increases. On the other hand a function $f \in L^p(\mu)$ is allowed to decay at "infinity" slower and slower as p increases. With these insights in mind, we should not in general expect $L^p(\mu) \subset L^q(\mu)$ or $L^q(\mu) \subset L^p(\mu)$. However, there are two notable exceptions. (1) If $\mu(\Omega) < \infty$, then there is no behavior at infinity to worry about and $L^q(\mu) \subset L^p(\mu)$ for all $q \geq p$ as is shown in Corollary 12.23 below. (2) If μ is counting measure, i.e. $\mu(A) = \#(A)$, then all functions in $L^p(\mu)$ for any p can not blow up on a set of positive measure, so there are no local singularities. In this case $L^p(\mu) \subset L^q(\mu)$ for all $q \geq p$, see Corollary 12.28 below.

Corollary 12.23. *If $\mu(\Omega) < \infty$ and $0 < p < q \leq \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \leq [\mu(\Omega)]^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q .$$

Proof. Take $a \in [1, \infty]$ such that

$$\frac{1}{p} = \frac{1}{a} + \frac{1}{q}, \text{ i.e. } a = \frac{pq}{q-p}.$$

Then by Theorem 12.14,

$$\|f\|_p = \|f \cdot 1\|_p \leq \|f\|_q \cdot \|1\|_a = \mu(\Omega)^{1/a} \|f\|_q = \mu(\Omega)^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q.$$

The reader may easily check this final formula is correct even when $q = \infty$ provided we interpret $1/p - 1/\infty$ to be $1/p$. ∎

The rest of this section may be skipped.

Example 12.24 (Power Inequalities). Let $a := (a_1, \ldots, a_n)$ with $a_i > 0$ for $i = 1, 2, \ldots, n$ and for $p \in \mathbb{R} \setminus \{0\}$, let

$$\|a\|_p := \left(\frac{1}{n} \sum_{i=1}^n a_i^p\right)^{1/p}.$$

Then by Corollary 12.23, $p \to \|a\|_p$ is increasing in p for $p > 0$. For $p = -q < 0$, we have

$$\|a\|_p := \left(\frac{1}{n} \sum_{i=1}^n a_i^{-q}\right)^{-1/q} = \left(\frac{1}{\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{a_i}\right)^q}\right)^{1/q} = \left\|\frac{1}{a}\right\|_q^{-1}$$

where $\frac{1}{a} := (1/a_1, \ldots, 1/a_n)$. So for $p < 0$, as p increases, $q = -p$ decreases, so that $\left\|\frac{1}{a}\right\|_q$ is decreasing and hence $\left\|\frac{1}{a}\right\|_q^{-1}$ is increasing. Hence we have shown that $p \to \|a\|_p$ is increasing for $p \in \mathbb{R} \setminus \{0\}$.

We now claim that $\lim_{p\to 0} \|a\|_p = \sqrt[n]{a_1 \ldots a_n}$. To prove this, write $a_i^p = e^{p \ln a_i} = 1 + p \ln a_i + O\left(p^2\right)$ for p near zero. Therefore,

$$\frac{1}{n} \sum_{i=1}^n a_i^p = 1 + p\frac{1}{n} \sum_{i=1}^n \ln a_i + O\left(p^2\right).$$

Hence it follows that

$$\lim_{p\to 0} \|a\|_p = \lim_{p\to 0} \left(\frac{1}{n} \sum_{i=1}^n a_i^p\right)^{1/p} = \lim_{p\to 0} \left(1 + p\frac{1}{n} \sum_{i=1}^n \ln a_i + O\left(p^2\right)\right)^{1/p}$$
$$= e^{\frac{1}{n}\sum_{i=1}^n \ln a_i} = \sqrt[n]{a_1 \ldots a_n}.$$

So if we now define $\|a\|_0 := \sqrt[n]{a_1 \ldots a_n}$, the map $p \in \mathbb{R} \to \|a\|_p \in (0, \infty)$ is continuous and increasing in p.

We will now show that $\lim_{p\to\infty} \|a\|_p = \max_i a_i =: M$ and $\lim_{p\to-\infty} \|a\|_p = \min_i a_i =: m$. Indeed, for $p > 0$,

$$\frac{1}{n} M^p \leq \frac{1}{n} \sum_{i=1}^n a_i^p \leq M^p$$

and therefore,

$$\left(\frac{1}{n}\right)^{1/p} M \leq \|a\|_p \leq M.$$

Since $\left(\frac{1}{n}\right)^{1/p} \to 1$ as $p \to \infty$, it follows that $\lim_{p\to\infty} \|a\|_p = M$. For $p = -q < 0$, we have

$$\lim_{p\to-\infty} \|a\|_p = \lim_{q\to\infty} \left(\frac{1}{\left\|\frac{1}{a}\right\|_q}\right) = \frac{1}{\max_i (1/a_i)} = \frac{1}{1/m} = m = \min_i a_i.$$

Conclusion. If we extend the definition of $\|a\|_p$ to $p = \infty$ and $p = -\infty$ by $\|a\|_\infty = \max_i a_i$ and $\|a\|_{-\infty} = \min_i a_i$, then $\bar{\mathbb{R}} \ni p \to \|a\|_p \in (0, \infty)$ is a continuous non-decreasing function of p.

Proposition 12.25. *Suppose that* $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ *and* $p_\lambda \in (p_0,p_1)$ *be defined by*

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1} \tag{12.20}$$

with the interpretation that $\lambda/p_1 = 0$ *if* $p_1 = \infty$.[3] *Then* $L^{p_\lambda} \subset L^{p_0} + L^{p_1}$, *i.e. every function* $f \in L^{p_\lambda}$ *may be written as* $f = g + h$ *with* $g \in L^{p_0}$ *and* $h \in L^{p_1}$. *For* $1 \le p_0 < p_1 \le \infty$ *and* $f \in L^{p_0} + L^{p_1}$ *let*

$$\|f\| := \inf\left\{\|g\|_{p_0} + \|h\|_{p_1} : f = g + h\right\}.$$

Then $(L^{p_0} + L^{p_1}, \|\cdot\|)$ *is a Banach space and the inclusion map from* L^{p_λ} *to* $L^{p_0} + L^{p_1}$ *is bounded; in fact* $\|f\| \le 2\|f\|_{p_\lambda}$ *for all* $f \in L^{p_\lambda}$.

Proof. Let $M > 0$, then the local singularities of f are contained in the set $E := \{|f| > M\}$ and the behavior of f at "infinity" is solely determined by f on E^c. Hence let $g = f1_E$ and $h = f1_{E^c}$ so that $f = g + h$. By our earlier discussion we expect that $g \in L^{p_0}$ and $h \in L^{p_1}$ and this is the case since,

$$\begin{aligned}\|g\|_{p_0}^{p_0} &= \int |f|^{p_0} 1_{|f|>M} = M^{p_0}\int \left|\frac{f}{M}\right|^{p_0} 1_{|f|>M} \\ &\le M^{p_0}\int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \le M^{p_0-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty\end{aligned}$$

and

$$\begin{aligned}\|h\|_{p_1}^{p_1} &= \left\|f1_{|f|\le M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\le M} = M^{p_1}\int \left|\frac{f}{M}\right|^{p_1} 1_{|f|\le M} \\ &\le M^{p_1}\int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\le M} \le M^{p_1-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty.\end{aligned}$$

Moreover this shows

$$\|f\| \le M^{1-p_\lambda/p_0}\|f\|_{p_\lambda}^{p_\lambda/p_0} + M^{1-p_\lambda/p_1}\|f\|_{p_\lambda}^{p_\lambda/p_1}.$$

Taking $M = \lambda\|f\|_{p_\lambda}$ then gives

$$\|f\| \le \left(\lambda^{1-p_\lambda/p_0} + \lambda^{1-p_\lambda/p_1}\right)\|f\|_{p_\lambda}$$

and then taking $\lambda = 1$ shows $\|f\| \le 2\|f\|_{p_\lambda}$. The proof that $(L^{p_0} + L^{p_1}, \|\cdot\|)$ is a Banach space is left as Exercise 12.10 to the reader. ∎

[3] A little algebra shows that λ may be computed in terms of p_0, p_λ and p_1 by

$$\lambda = \frac{p_0}{p_\lambda}\cdot\frac{p_1 - p_\lambda}{p_1 - p_0}.$$

Corollary 12.26 (Interpolation of L^p – norms). *Suppose that* $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ *and* $p_\lambda \in (p_0,p_1)$ *be defined as in Eq. (12.20), then* $L^{p_0} \cap L^{p_1} \subset L^{p_\lambda}$ *and*

$$\|f\|_{p_\lambda} \le \|f\|_{p_0}^{\lambda}\|f\|_{p_1}^{1-\lambda}. \tag{12.21}$$

Further assume $1 \le p_0 < p_\lambda < p_1 \le \infty$, *and for* $f \in L^{p_0} \cap L^{p_1}$ *let*

$$\|f\| := \|f\|_{p_0} + \|f\|_{p_1}.$$

Then $(L^{p_0} \cap L^{p_1}, \|\cdot\|)$ *is a Banach space and the inclusion map of* $L^{p_0} \cap L^{p_1}$ *into* L^{p_λ} *is bounded, in fact*

$$\|f\|_{p_\lambda} \le \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{12.22}$$

The heuristic explanation of this corollary is that if $f \in L^{p_0} \cap L^{p_1}$, then f has local singularities no worse than an L^{p_1} function and behavior at infinity no worse than an L^{p_0} function. Hence $f \in L^{p_\lambda}$ for any p_λ between p_0 and p_1.

Proof. Let λ be determined as above, $a = p_0/\lambda$ and $b = p_1/(1-\lambda)$, then by Theorem 12.14,

$$\|f\|_{p_\lambda} = \left\||f|^{\lambda}|f|^{1-\lambda}\right\|_{p_\lambda} \le \left\||f|^{\lambda}\right\|_a \left\||f|^{1-\lambda}\right\|_b = \|f\|_{p_0}^{\lambda}\|f\|_{p_1}^{1-\lambda}.$$

It is easily checked that $\|\cdot\|$ is a norm on $L^{p_0} \cap L^{p_1}$. To show this space is complete, suppose that $\{f_n\} \subset L^{p_0} \cap L^{p_1}$ is a $\|\cdot\|$ – Cauchy sequence. Then $\{f_n\}$ is both L^{p_0} and L^{p_1} – Cauchy. Hence there exist $f \in L^{p_0}$ and $g \in L^{p_1}$ such that $\lim_{n\to\infty}\|f - f_n\|_{p_0} = 0$ and $\lim_{n\to\infty}\|g - f_n\|_{p_\lambda} = 0$. By Chebyshev's inequality (Lemma 12.4) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 12.7, $f = g$ a.e. It now is clear that $\lim_{n\to\infty}\|f - f_n\| = 0$. The estimate in Eq. (12.22) is left as Exercise 12.9 to the reader. ∎

Remark 12.27. Combining Proposition 12.25 and Corollary 12.26 gives

$$L^{p_0} \cap L^{p_1} \subset L^{p_\lambda} \subset L^{p_0} + L^{p_1}$$

for $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0,p_1)$ as in Eq. (12.20).

Corollary 12.28. *Suppose now that* μ *is counting measure on* Ω. *Then* $L^p(\mu) \subset L^q(\mu)$ *for all* $0 < p < q \le \infty$ *and* $\|f\|_q \le \|f\|_p$.

Proof. Suppose that $0 < p < q = \infty$, then

$$\|f\|_\infty^p = \sup\{|f(x)|^p : x \in \Omega\} \le \sum_{x\in\Omega}|f(x)|^p = \|f\|_p^p,$$

i.e. $\|f\|_\infty \le \|f\|_p$ for all $0 < p < \infty$. For $0 < p \le q \le \infty$, apply Corollary 12.26 with $p_0 = p$ and $p_1 = \infty$ to find

$$\|f\|_q \le \|f\|_p^{p/q}\|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q}\|f\|_p^{1-p/q} = \|f\|_p.$$

∎

12.5.1 Summary:

1. $L^{p_0} \cap L^{p_1} \subset L^q \subset L^{p_0} + L^{p_1}$ for any $q \in (p_0, p_1)$.
2. If $p \le q$, then $\ell^p \subset \ell^q$ and $\|f\|_q \le \|f\|_p$.
3. Since $\mu(|f| > \varepsilon) \le \varepsilon^{-p} \|f\|_p^p$, L^p – convergence implies L^0 – convergence.
4. L^0 – convergence implies almost everywhere convergence for some subsequence.
5. If $\mu(\Omega) < \infty$ then almost everywhere convergence implies uniform convergence off certain sets of small measure and in particular we have L^0 – convergence.
6. If $\mu(\Omega) < \infty$, then $L^q \subset L^p$ for all $p \le q$ and L^q – convergence implies L^p – convergence.

12.6 Uniform Integrability

This section will address the question as to what extra conditions are needed in order that an L^0 – convergent sequence is L^p – convergent. This will lead us to the notion of uniform integrability. To simplify matters a bit here, it will be assumed that $(\Omega, \mathcal{B}, \mu)$ is a finite measure space for this section.

Notation 12.29 *For $f \in L^1(\mu)$ and $E \in \mathcal{B}$, let*

$$\mu(f : E) := \int_E f d\mu.$$

and more generally if $A, B \in \mathcal{B}$ let

$$\mu(f : A, B) := \int_{A \cap B} f d\mu.$$

When μ is a probability measure, we will often write $\mathbb{E}[f : E]$ for $\mu(f : E)$ and $\mathbb{E}[f : A, B]$ for $\mu(f : A, B)$.

Definition 12.30. *A collection of functions, $\Lambda \subset L^1(\mu)$ is said to be **uniformly integrable** if,*

$$\lim_{a \to \infty} \sup_{f \in \Lambda} \mu(|f| : |f| \ge a) = 0. \tag{12.23}$$

The condition in Eq. (12.23) implies $\sup_{f \in \Lambda} \|f\|_1 < \infty$.[4] Indeed, choose a sufficiently large so that $\sup_{f \in \Lambda} \mu(|f| : |f| \ge a) \le 1$, then for $f \in \Lambda$

[4] This is not necessarily the case if $\mu(\Omega) = \infty$. Indeed, if $\Omega = \mathbb{R}$ and $\mu = m$ is Lebesgue measure, the sequences of functions, $\left\{f_n := 1_{[-n,n]}\right\}_{n=1}^{\infty}$ are uniformly integrable but not bounded in $L^1(m)$.

$$\|f\|_1 = \mu(|f| : |f| \ge a) + \mu(|f| : |f| < a) \le 1 + a\mu(\Omega).$$

Let us also note that if $\Lambda = \{f\}$ with $f \in L^1(\mu)$, then Λ is uniformly integrable. Indeed, $\lim_{a \to \infty} \mu(|f| : |f| \ge a) = 0$ by the dominated convergence theorem.

Exercise 12.6. Suppose A is an index set, $\{f_\alpha\}_{\alpha \in A}$ and $\{g_\alpha\}_{\alpha \in A}$ are two collections of random variables. If $\{g_\alpha\}_{\alpha \in A}$ is uniformly integrable and $|f_\alpha| \le |g_\alpha|$ for all $\alpha \in A$, show $\{f_\alpha\}_{\alpha \in A}$ is uniformly integrable as well.

Solution to Exercise (12.6). For $a > 0$ we have

$$\mathbb{E}[|f_\alpha| : |f_\alpha| \ge a] \le \mathbb{E}[|g_\alpha| : |f_\alpha| \ge a] \le \mathbb{E}[|g_\alpha| : |g_\alpha| \ge a].$$

Therefore,

$$\lim_{a \to \infty} \sup_\alpha \mathbb{E}[|f_\alpha| : |f_\alpha| \ge a] \le \lim_{a \to \infty} \sup_\alpha \mathbb{E}[|g_\alpha| : |g_\alpha| \ge a] = 0.$$

Definition 12.31. *A collection of functions, $\Lambda \subset L^1(\mu)$ is said to be **uniformly absolutely continuous** if for all $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\sup_{f \in \Lambda} \mu(|f| : E) < \varepsilon \text{ whenever } \mu(E) < \delta. \tag{12.24}$$

Remark 12.32. It is not in general true that if $\{f_n\} \subset L^1(\mu)$ is uniformly absolutely continuous implies $\sup_n \|f_n\|_1 < \infty$. For example take $\Omega = \{*\}$ and $\mu(\{*\}) = 1$. Let $f_n(*) = n$. Since for $\delta < 1$ a set $E \subset \Omega$ such that $\mu(E) < \delta$ is in fact the empty set and hence $\{f_n\}_{n=1}^\infty$ is uniformly absolutely continuous. However, for finite measure spaces without "atoms", for every $\delta > 0$ we may find a finite partition of Ω by sets $\{E_\ell\}_{\ell=1}^k$ with $\mu(E_\ell) < \delta$. If Eq. (12.24) holds with $\varepsilon = 1$, then

$$\mu(|f_n|) = \sum_{\ell=1}^k \mu(|f_n| : E_\ell) \le k$$

showing that $\mu(|f_n|) \le k$ for all n.

Lemma 12.33 (This lemma may be skipped.). *For any $g \in L^1(\mu)$, $\Lambda = \{g\}$ is uniformly absolutely continuous.*

Proof. First Proof. If the Lemma is false, there would exist $\varepsilon > 0$ and sets E_n such that $\mu(E_n) \to 0$ while $\mu(|g| : E_n) \ge \varepsilon$ for all n. Since $|1_{E_n} g| \le |g| \in L^1$ and for any $\delta > 0$, $\mu(1_{E_n}|g| > \delta) \le \mu(E_n) \to 0$ as $n \to \infty$, the dominated convergence theorem of Corollary 12.8 implies $\lim_{n \to \infty} \mu(|g| : E_n) = 0$. This contradicts $\mu(|g| : E_n) \ge \varepsilon$ for all n and the proof is complete.

Second Proof. Let $\varphi = \sum_{i=1}^n c_i 1_{B_i}$ be a simple function such that $\|g - \varphi\|_1 < \varepsilon/2$. Then

$$\mu(|g| : E) \le \mu(|\varphi| : E) + \mu(|g-\varphi| : E)$$
$$\le \sum_{i=1}^{n} |c_i| \mu(E \cap B_i) + \|g-\varphi\|_1 \le \left(\sum_{i=1}^{n} |c_i|\right) \mu(E) + \varepsilon/2.$$

This shows $\mu(|g| : E) < \varepsilon$ provided that $\mu(E) < \varepsilon \left(2 \sum_{i=1}^{n} |c_i|\right)^{-1}$. ■

Proposition 12.34. *A subset $\Lambda \subset L^1(\mu)$ is uniformly integrable iff $\Lambda \subset L^1(\mu)$ is bounded and uniformly absolutely continuous.*

Proof. ($\Longrightarrow$) We have already seen that uniformly integrable subsets, Λ, are bounded in $L^1(\mu)$. Moreover, for $f \in \Lambda$, and $E \in \mathcal{B}$,

$$\mu(|f| : E) = \mu(|f| : |f| \ge M, E) + \mu(|f| : |f| < M, E)$$
$$\le \sup_n \mu(|f| : |f| \ge M) + M\mu(E).$$

So given $\varepsilon > 0$ choose M so large that $\sup_{f\in\Lambda} \mu(|f| : |f| \ge M) < \varepsilon/2$ and then take $\delta = \frac{\varepsilon}{2M}$ to verify that Λ is uniformly absolutely continuous.

($\Longleftarrow$) Let $K := \sup_{f\in\Lambda} \|f\|_1 < \infty$. Then for $f \in \Lambda$, we have

$$\mu(|f| \ge a) \le \|f\|_1 / a \le K/a \text{ for all } a > 0.$$

Hence given $\varepsilon > 0$ and $\delta > 0$ as in the definition of uniform absolute continuity, we may choose $a = K/\delta$ in which case

$$\sup_{f\in\Lambda} \mu(|f| : |f| \ge a) < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\lim_{a\to\infty} \sup_{f\in\Lambda} \mu(|f| : |f| \ge a) = 0$ as desired. ■

Corollary 12.35. *Suppose $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are two uniformly integrable collections of functions, then $\{f_\alpha + g_\alpha\}_{\alpha\in A}$ is also uniformly integrable.*

Proof. By Proposition 12.34, $\{f_\alpha\}_{\alpha\in A}$ and $\{g_\alpha\}_{\alpha\in A}$ are both bounded in $L^1(\mu)$ and are both uniformly absolutely continuous. Since $\|f_\alpha + g_\alpha\|_1 \le \|f_\alpha\|_1 + \|g_\alpha\|_1$ it follows that $\{f_\alpha + g_\alpha\}_{\alpha\in A}$ is bounded in $L^1(\mu)$ as well. Moreover, for $\varepsilon > 0$ we may choose $\delta > 0$ such that $\mu(|f_\alpha| : E) < \varepsilon$ and $\mu(|g_\alpha| : E) < \varepsilon$ whenever $\mu(E) < \delta$. For this choice of ε and δ, we then have

$$\mu(|f_\alpha + g_\alpha| : E) \le \mu(|f_\alpha| + |g_\alpha| : E) < 2\varepsilon \text{ whenever } \mu(E) < \delta,$$

showing $\{f_\alpha + g_\alpha\}_{\alpha\in A}$ uniformly absolutely continuous. Another application of Proposition 12.34 completes the proof. ■

Exercise 12.7 (Problem 5 on p. 196 of Resnick.). Suppose that $\{X_n\}_{n=1}^\infty$ is a sequence of integrable and i.i.d random variables. Then $\left\{\frac{S_n}{n}\right\}_{n=1}^\infty$ is uniformly integrable.

Theorem 12.36 (Vitali Convergence Theorem). *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space, $\Lambda := \{f_n\}_{n=1}^\infty$ be a sequence of functions in $L^1(\mu)$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then $f \in L^1(\mu)$ and $\|f - f_n\|_1 \to 0$ as $n \to \infty$ iff $f_n \to f$ in μ measure and Λ is uniformly integrable.*

Proof. ($\Longleftarrow$) If $f_n \to f$ in μ measure and $\Lambda = \{f_n\}_{n=1}^\infty$ is uniformly integrable then we know $M := \sup_n \|f_n\|_1 < \infty$. Hence and application of Fatou's lemma, see Exercise 12.1,

$$\int_\Omega |f| \, d\mu \le \liminf_{n\to\infty} \int_\Omega |f_n| \, d\mu \le M < \infty,$$

i.e. $f \in L^1(\mu)$. One now easily checks that $\Lambda_0 := \{f - f_n\}_{n=1}^\infty$ is bounded in $L^1(\mu)$ and (using Lemma 12.33 and Proposition 12.34) Λ_0 is uniformly absolutely continuous and hence Λ_0 is uniformly integrable. Therefore,

$$\|f - f_n\|_1 = \mu(|f - f_n| : |f - f_n| \ge a) + \mu(|f - f_n| : |f - f_n| < a)$$
$$\le \varepsilon(a) + \int_\Omega 1_{|f-f_n|<a} |f - f_n| \, d\mu \tag{12.25}$$

where

$$\varepsilon(a) := \sup_m \mu(|f - f_m| : |f - f_m| \ge a) \to 0 \text{ as } a \to \infty.$$

Since $1_{|f-f_n|<a} |f - f_n| \le a \in L^1(\mu)$ and

$$\mu\left(1_{|f-f_n|<a} |f - f_n| > \varepsilon\right) \le \mu(|f - f_n| > \varepsilon) \to 0 \text{ as } n \to \infty,$$

we may pass to the limit in Eq. (12.25), with the aid of the dominated convergence theorem (see Corollary 12.8), to find

$$\limsup_{n\to\infty} \|f - f_n\|_1 \le \varepsilon(a) \to 0 \text{ as } a \to \infty.$$

($\Longrightarrow$) If $f_n \to f$ in $L^1(\mu)$, then by Chebyschev's inequality it follows that $f_n \to f$ in μ – measure. Since convergent sequences are bounded, to show Λ is uniformly integrable it suffices to shows Λ is uniformly absolutely continuous. Now for $E \in \mathcal{B}$ and $n \in \mathbb{N}$,

$$\mu(|f_n| : E) \le \mu(|f - f_n| : E) + \mu(|f| : E) \le \|f - f_n\|_1 + \mu(|f| : E).$$

Let $\varepsilon_N := \sup_{n>N} \|f - f_n\|_1$, then $\varepsilon_N \downarrow 0$ as $N \uparrow \infty$ and

$$\sup_n \mu(|f_n| : E) \le \sup_{n\le N} \mu(|f_n| : E) \vee (\varepsilon_N + \mu(|f| : E)) \le \varepsilon_N + \mu(g_N : E), \tag{12.26}$$

where $g_N = |f| + \sum_{n=1}^N |f_n| \in L^1$. Given $\varepsilon > 0$ fix N large so that $\varepsilon_N < \varepsilon/2$ and then choose $\delta > 0$ (by Lemma 12.33) such that $\mu(g_N : E) < \varepsilon$ if $\mu(E) < \delta$. It then follows from Eq. (12.26) that

$$\sup_n \mu(|f_n| : E) < \varepsilon/2 + \varepsilon/2 = \varepsilon \text{ when } \mu(E) < \delta.$$

■

Example 12.37. Let $\Omega = [0,1]$, $\mathcal{B} = \mathcal{B}_{[0,1]}$ and $P = m$ be Lebesgue measure on $\mathcal{B}$. Then the collection of functions, $f_\varepsilon(x) := \frac{2}{\varepsilon}(1 - x/\varepsilon) \vee 0$ for $\varepsilon \in (0,1)$ is bounded in $L^1(P)$, $f_\varepsilon \to 0$ a.e. as $\varepsilon \downarrow 0$ but

$$0 = \int_\Omega \lim_{\varepsilon\downarrow 0} f_\varepsilon dP \ne \lim_{\varepsilon\downarrow 0} \int_\Omega f_\varepsilon dP = 1.$$

This is a typical example of a bounded and pointwise convergent sequence in L^1 which is not uniformly integrable.

Example 12.38. Let $\Omega = [0,1]$, P be Lebesgue measure on $\mathcal{B} = \mathcal{B}_{[0,1]}$, and for $\varepsilon \in (0,1)$ let $a_\varepsilon > 0$ with $\lim_{\varepsilon\downarrow 0} a_\varepsilon = \infty$ and let $f_\varepsilon := a_\varepsilon 1_{[0,\varepsilon]}$. Then $\mathbb{E}f_\varepsilon = \varepsilon a_\varepsilon$ and so $\sup_{\varepsilon>0} \|f_\varepsilon\|_1 =: K < \infty$ iff $\varepsilon a_\varepsilon \le K$ for all ε. Since

$$\sup_\varepsilon \mathbb{E}[f_\varepsilon : f_\varepsilon \ge M] = \sup_\varepsilon [\varepsilon a_\varepsilon \cdot 1_{a_\varepsilon \ge M}],$$

if $\{f_\varepsilon\}$ is uniformly integrable and $\delta > 0$ is given, for large M we have $\varepsilon a_\varepsilon \le \delta$ for ε small enough so that $a_\varepsilon \ge M$. From this we conclude that $\limsup_{\varepsilon\downarrow 0}(\varepsilon a_\varepsilon) \le \delta$ and since $\delta > 0$ was arbitrary, $\lim_{\varepsilon\downarrow 0} \varepsilon a_\varepsilon = 0$ if $\{f_\varepsilon\}$ is uniformly integrable. By reversing these steps one sees the converse is also true.

Alternatively. No matter how $a_\varepsilon > 0$ is chosen, $\lim_{\varepsilon\downarrow 0} f_\varepsilon = 0$ a.s.. So from Theorem 12.36, if $\{f_\varepsilon\}$ is uniformly integrable we would have to have

$$\lim_{\varepsilon\downarrow 0}(\varepsilon a_\varepsilon) = \lim_{\varepsilon\downarrow 0} \mathbb{E}f_\varepsilon = \mathbb{E}0 = 0.$$

Corollary 12.39. *Let $(\Omega, \mathcal{B}, \mu)$ be a finite measure space, $p \in [1,\infty)$, $\{f_n\}_{n=1}^\infty$ be a sequence of functions in $L^p(\mu)$, and $f : \Omega \to \mathbb{C}$ be a measurable function. Then $f \in L^p(\mu)$ and $\|f - f_n\|_p \to 0$ as $n \to \infty$ iff $f_n \to f$ in μ measure and $\Lambda := \{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable.*

Proof. ($\Longleftarrow$) Suppose that $f_n \to f$ in μ measure and $\Lambda := \{|f_n|^p\}_{n=1}^\infty$ is uniformly integrable. By Corollary 12.9, $|f_n|^p \xrightarrow{\mu} |f|^p$ in μ – measure, and $h_n := |f - f_n|^p \xrightarrow{\mu} 0$, and by Theorem 12.36, $|f|^p \in L^1(\mu)$ and $|f_n|^p \to |f|^p$ in $L^1(\mu)$. Since

$$h_n := |f - f_n|^p \le (|f| + |f_n|)^p \le 2^{p-1}(|f|^p + |f_n|^p) =: g_n \in L^1(\mu)$$

with $g_n \to g := 2^{p-1}|f|^p$ in $L^1(\mu)$, the dominated convergence theorem in Corollary 12.8, implies

$$\|f - f_n\|_p^p = \int_\Omega |f - f_n|^p d\mu = \int_\Omega h_n d\mu \to 0 \text{ as } n \to \infty.$$

($\Longrightarrow$) Suppose $f \in L^p$ and $f_n \to f$ in L^p. Again $f_n \to f$ in μ – measure by Lemma 12.4. Let

$$h_n := ||f_n|^p - |f|^p| \le |f_n|^p + |f|^p =: g_n \in L^1$$

and $g := 2|f|^p \in L^1$. Then $g_n \xrightarrow{\mu} g$, $h_n \xrightarrow{\mu} 0$ and $\int g_n d\mu \to \int g d\mu$. Therefore by the dominated convergence theorem in Corollary 12.8, $\lim_{n\to\infty} \int h_n \, d\mu = 0$, i.e. $|f_n|^p \to |f|^p$ in $L^1(\mu)$.[5] Hence it follows from Theorem 12.36 that Λ is uniformly integrable. ■

The following Lemma gives a concrete necessary and sufficient conditions for verifying a sequence of functions is uniformly integrable.

Lemma 12.40. *Suppose that $\mu(\Omega) < \infty$, and $\Lambda \subset L^0(\Omega)$ is a collection of functions.*

1. *If there exists a non decreasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{x\to\infty} \varphi(x)/x = \infty$ and*
$$K := \sup_{f\in\Lambda} \mu(\varphi(|f|)) < \infty \tag{12.27}$$
then Λ is uniformly integrable.
2. *Conversely if Λ is uniformly integrable, there exists a non-decreasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi(0) = 0$, $\lim_{x\to\infty} \varphi(x)/x = \infty$ and Eq. (12.27) is valid.*

[5] Here is an alternative proof. By the mean value theorem,

$$||f|^p - |f_n|^p| \le p(\max(|f|, |f_n|))^{p-1} ||f| - |f_n|| \le p(|f| + |f_n|)^{p-1} ||f| - |f_n||$$

and therefore by Hölder's inequality,

$$\begin{aligned}\int ||f|^p - |f_n|^p| \, d\mu &\le p \int (|f| + |f_n|)^{p-1} ||f| - |f_n|| \, d\mu \le p \int (|f| + |f_n|)^{p-1} |f - f_n| \, d\mu \\ &\le p\|f - f_n\|_p \|(|f| + |f_n|)^{p-1}\|_q = p\| \, |f| + |f_n|\|_p^{p/q} \|f - f_n\|_p \\ &\le p(\|f\|_p + \|f_n\|_p)^{p/q} \|f - f_n\|_p\end{aligned}$$

where $q := p/(p-1)$. This shows that $\int ||f|^p - |f_n|^p| \, d\mu \to 0$ as $n \to \infty$.

A typical example for φ in item 1. is $\varphi(x) = x^p$ for some $p > 1$.

Proof. 1. Let φ be as in item 1. above and set $\varepsilon_a := \sup_{x \geq a} \frac{x}{\varphi(x)} \to 0$ as $a \to \infty$ by assumption. Then for $f \in \Lambda$

$$\begin{aligned}\mu(|f| : |f| \geq a) &= \mu\left(\frac{|f|}{\varphi(|f|)}\varphi(|f|) : |f| \geq a\right) \leq \mu(\varphi(|f|) : |f| \geq a)\varepsilon_a \\ &\leq \mu(\varphi(|f|))\varepsilon_a \leq K\varepsilon_a\end{aligned}$$

and hence

$$\lim_{a\to\infty} \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\geq a}\right) \leq \lim_{a\to\infty} K\varepsilon_a = 0.$$

2. By assumption, $\varepsilon_a := \sup_{f\in\Lambda} \mu\left(|f| 1_{|f|\geq a}\right) \to 0$ as $a \to \infty$. Therefore we may choose $a_n \uparrow \infty$ such that

$$\sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty$$

where by convention $a_0 := 0$. Now define φ so that $\varphi(0) = 0$ and

$$\varphi'(x) = \sum_{n=0}^{\infty} (n+1)\, 1_{(a_n, a_{n+1}]}(x),$$

i.e.

$$\varphi(x) = \int_0^x \varphi'(y)dy = \sum_{n=0}^{\infty} (n+1)\left(x \wedge a_{n+1} - x \wedge a_n\right).$$

By construction φ is continuous, $\varphi(0) = 0$, $\varphi'(x)$ is increasing (so φ is convex) and $\varphi'(x) \geq (n+1)$ for $x \geq a_n$. In particular

$$\frac{\varphi(x)}{x} \geq \frac{\varphi(a_n) + (n+1)x}{x} \geq n+1 \text{ for } x \geq a_n$$

from which we conclude $\lim_{x\to\infty} \varphi(x)/x = \infty$. We also have $\varphi'(x) \leq (n+1)$ on $[0, a_{n+1}]$ and therefore

$$\varphi(x) \leq (n+1)x \text{ for } x \leq a_{n+1}.$$

So for $f \in \Lambda$,

$$\begin{aligned}\mu(\varphi(|f|)) &= \sum_{n=0}^{\infty} \mu\left(\varphi(|f|)1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{(a_n,a_{n+1}]}(|f|)\right) \\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{|f|\geq a_n}\right) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n}\end{aligned}$$

and hence

$$\sup_{f\in\Lambda} \mu(\varphi(|f|)) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty.$$

■

12.7 Exercises

Exercise 12.8. Let $f \in L^p \cap L^\infty$ for some $p < \infty$. Show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$. If we further assume $\mu(X) < \infty$, show $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$ for all measurable functions $f : X \to \mathbb{C}$. In particular, $f \in L^\infty$ iff $\lim_{q\to\infty} \|f\|_q < \infty$. **Hints:** Use Corollary 12.26 to show $\limsup_{q\to\infty} \|f\|_q \leq \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \geq \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

Exercise 12.9. Prove Eq. (12.22) in Corollary 12.26. (Part of Folland 6.3 on p. 186.) **Hint:** Use the inequality, with $a, b \geq 1$ with $a^{-1} + b^{-1} = 1$ chosen appropriately,

$$st \leq \frac{s^a}{a} + \frac{t^b}{b}$$

applied to the right side of Eq. (12.21).

Exercise 12.10. Complete the proof of Proposition 12.25 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space.

12.8 Appendix: Convex Functions

Reference; see the appendix (page 500) of Revuz and Yor.

Definition 12.41. *Given any function, $\varphi : (a,b) \to \mathbb{R}$, we say that φ is **convex** if for all $a < x_0 \leq x_1 < b$ and $t \in [0,1]$,*

$$\varphi(x_t) \leq h_t := (1-t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in [0,1], \tag{12.28}$$

where

$$x_t := x_0 + t(x_1 - x_0) = (1-t)x_0 + tx_1, \tag{12.29}$$

see Figure 12.1 below.

Lemma 12.42. *Let $\varphi : (a,b) \to \mathbb{R}$ be a function and*

$$F(x_0, x_1) := \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} \text{ for } a < x_0 < x_1 < b.$$

Then the following are equivalent;

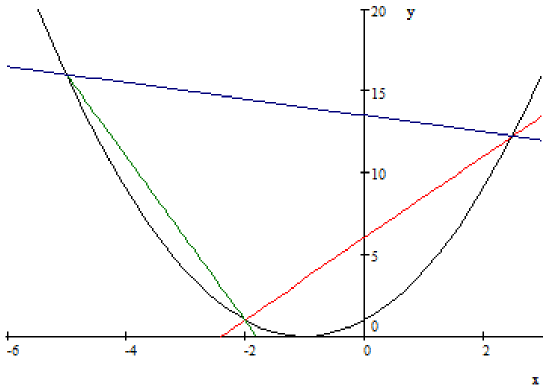


Fig. 12.1. A convex function along with three cords corresponding to $x_0 = -5$ and $x_1 = -2$ $x_0 = -2$ and $x_1 = 5/2$, and $x_0 = -5$ and $x_1 = 5/2$ with slopes, $m_1 = -15/3$, $m_2 = 15/6$ and $m_3 = -1/2$ respectively. Notice that $m_1 \le m_3 \le m_2$.

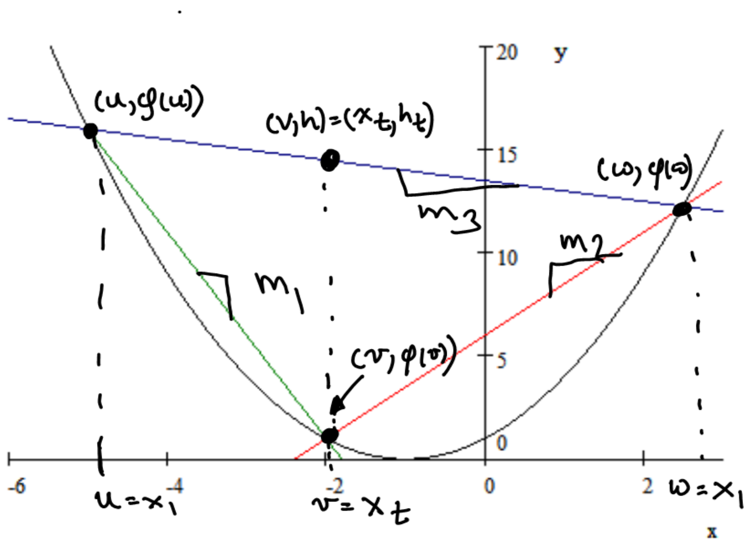


Fig. 12.2. A convex function with three cords. Notice the slope relationships; $m_1 \le m_3 \le m_2$.

1. φ *is convex,*
2. $F(x_0, x_1)$ *is non-decreasing in* x_0 *for all* $a < x_0 < x_1 < b$, *and*
3. $F(x_0, x_1)$ *is non-decreasing in* x_1 *for all* $a < x_0 < x_1 < b$.

Proof. Let x_t and h_t be as in Eq. (12.28), then (x_t, h_t) is on the line segment joining $(x_0, \varphi(x_0))$ to $(x_1, \varphi(x_1))$ and the statement that φ is convex is then equivalent to the assertion that $\varphi(x_t) \le h_t$ for all $0 \le t \le 1$. Since (x_t, h_t) lies on a straight line we always have the following three slopes are equal;

$$\frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t}.$$

In light of this identity, it is now clear that the convexity of φ is equivalent to either,

$$F(x_0, x_t) = \frac{\varphi(x_t) - \varphi(x_0)}{x_t - x_0} \le \frac{h_t - \varphi(x_0)}{x_t - x_0} = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = F(x_0, x_1)$$

or

$$F(x_0, x_1) = \frac{\varphi(x_1) - \varphi(x_0)}{x_1 - x_0} = \frac{\varphi(x_1) - h_t}{x_1 - x_t} \le \frac{\varphi(x_1) - \varphi(x_t)}{x_1 - x_t} = F(x_t, x_1)$$

holding for all $x_0 < x_t < x_1$. ■

Lemma 12.43 (A generalized FTC). *If* $\varphi \in PC^1((a,b) \to \mathbb{R})$[6], *then for all* $a < x < y < b$,

$$\varphi(y) - \varphi(x) = \int_x^y \varphi'(t)\, dt.$$

Proof. Let $b_1, \dots, b_{l-1}$ be the points of non-differentiability of φ in (x, y) and set $b_0 = x$ and $b_l = y$. Then

$$\varphi(y) - \varphi(x) = \sum_{k=1}^{l} [\varphi(b_k) - \varphi(b_{k-1})]$$
$$= \sum_{k=1}^{l} \int_{b_{k-1}}^{b_k} \varphi'(t)\, dt = \int_x^y \varphi'(t)\, dt.$$

■

Figure 12.3 below serves as motivation for the following elementary lemma on convex functions.

[6] PC^1 denotes the space of piecewise C^1 – functions, i.e. $\varphi \in PC^1((a,b) \to \mathbb{R})$ means the φ is continuous and there are a finite number of points,

$$\{a = a_0 < a_1 < a_2 < \dots < a_{n-1} < a_n = b\},$$

such that $\varphi|_{[a_{j-1}, a_j] \cap (a,b)}$ is C^1 for all $j = 1, 2, \dots, n$.

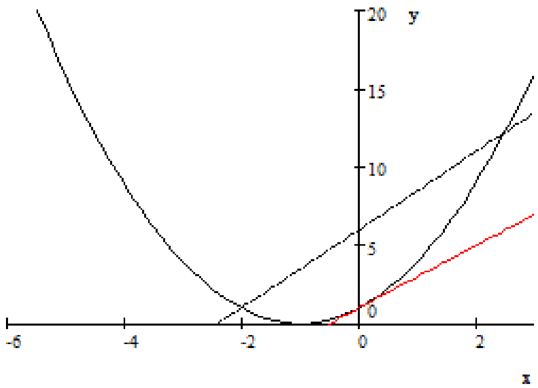


Fig. 12.3. A convex function, φ, along with a cord and a tangent line. Notice that the tangent line is always below φ and the cord lies above φ between the points of intersection of the cord with the graph of φ.

Lemma 12.44 (Convex Functions). *Let $\varphi \in PC^1((a,b) \to \mathbb{R})$ and for $x \in (a,b)$, let*

$$\varphi'(x+) := \lim_{h \downarrow 0} \frac{\varphi(x+h) - \varphi(x)}{h} \text{ and}$$
$$\varphi'(x-) := \lim_{h \uparrow 0} \frac{\varphi(x+h) - \varphi(x)}{h}.$$

(Of course, $\varphi'(x\pm) = \varphi'(x)$ at points $x \in (a,b)$ where φ is differentiable.)

1. *If $\varphi'(x) \le \varphi'(y)$ for all $a < x < y < b$ with x and y be points where φ is differentiable, then for any $x_0 \in (a,b)$, we have $\varphi'(x_0-) \le \varphi'(x_0+)$ and for $m \in (\varphi'(x_0-), \varphi'(x_0+))$ we have,*
$$\varphi(x_0) + m(x - x_0) \le \varphi(x) \ \forall\ x_0, x \in (a,b). \tag{12.30}$$
2. *If $\varphi \in PC^2((a,b) \to \mathbb{R})$[7] with $\varphi''(x) \ge 0$ for almost all $x \in (a,b)$, then Eq. (12.30) holds with $m = \varphi'(x_0)$.*
3. *If either of the hypothesis in items 1. and 2. above hold then φ is convex.*

(This lemma applies to the functions, $e^{\lambda x}$ for all $\lambda \in \mathbb{R}$, $|x|^\alpha$ for $\alpha > 1$, and $-\ln x$ to name a few examples. See Appendix 12.8 below for much more on convex functions.)

[7] PC^2 denotes the space of piecewise C^2 – functions, i.e. $\varphi \in PC^2((a,b) \to \mathbb{R})$ means the φ is C^1 and there are a finite number of points,
$$\{a = a_0 < a_1 < a_2 < \cdots < a_{n-1} < a_n = b\},$$
such that $\varphi|_{[a_{j-1},a_j] \cap (a,b)}$ is C^2 for all $j = 1, 2, \dots, n$.

Proof. 1. If x_0 is a point where φ is not differentiable and $h > 0$ is small, by the mean value theorem, for all $h > 0$ small, there exists $c_+(h) \in (x_0, x_0 + h)$ and $c_-(h) \in (x_0 - h, x_0)$ such that

$$\frac{\varphi(x_0 - h) - \varphi(x_0)}{-h} = \varphi'(c_-(h)) \le \varphi'(c_+(h)) = \frac{\varphi(x_0 + h) - \varphi(x_0)}{h}.$$

Letting $h \downarrow 0$ in this equation shows $\varphi'(x_0-) \le \varphi'(x_0+)$. Furthermore if $x < x_0 < y$ with x and y being points of differentiability of φ, the for small $h > 0$,
$$\varphi'(x) \le \varphi'(c_-(h)) \le \varphi'(c_+(h)) \le \varphi'(y).$$
Letting $h \downarrow 0$ in these inequalities shows,

$$\varphi'(x) \le \varphi'(x_0-) \le \varphi'(x_0+) \le \varphi'(y). \tag{12.31}$$

Now let $m \in (\varphi'(x_0-), \varphi'(x_0+))$. By the fundamental theorem of calculus in Lemma 12.43 and making use of Eq. (12.31), if $x > x_0$ then

$$\varphi(x) - \varphi(x_0) = \int_{x_0}^{x} \varphi'(t)\,dt \ge \int_{x_0}^{x} m\ dt = m(x - x_0)$$

and if $x < x_0$, then

$$\varphi(x_0) - \varphi(x) = \int_{x}^{x_0} \varphi'(t)\,dt \le \int_{x}^{x_0} m\ dt = m(x_0 - x).$$

These two equations implies Eq. (12.30) holds.

2. Notice that $\varphi' \in PC^1((a,b))$ and therefore,

$$\varphi'(y) - \varphi'(x) = \int_x^y \varphi''(t)\,dt \ge 0 \text{ for all } a < x \le y < b$$

which shows that item 1. may be used.

Alternatively; by Taylor's theorem with integral remainder (see Eq. (7.54) with $F = \varphi$, $a = x_0$, and $b = x$) implies

$$\varphi(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0) + (x - x_0)^2 \int_0^1 \varphi''(x_0 + \tau(x - x_0))(1 - \tau)\,d\tau$$
$$\ge \varphi(x_0) + \varphi'(x_0)(x - x_0).$$

3. For any $\xi\in(a,b)$, let $h_\xi(x) := \varphi(x_0)+\varphi'(x_0)(x-x_0)$. By Eq. (12.30) we know that $h_\xi(x)\le\varphi(x)$ for all $\xi,x\in(a,b)$ with equality when $\xi=x$ and therefore,

$$\varphi(x)=\sup_{\xi\in(a,b)} h_\xi(x).$$

Since h_ξ is an affine function for each $\xi\in(a,b)$, it follows that

$$h_\xi(x_t)=(1-t)h_\xi(x_0)+th_\xi(x_1)\le(1-t)\varphi(x_0)+t\varphi(x_1)$$

for all $t\in[0,1]$. Thus we may conclude that

$$\varphi(x_t)=\sup_{\xi\in(a,b)} h_\xi(x_t)\le(1-t)\varphi(x_0)+t\varphi(x_1)$$

as desired.

*For fun, here are three more proofs of Eq. (12.28) under the hypothesis of item 2. Clearly these proofs may be omitted.

3a. By Lemma 12.42 below it suffices to show either

$$\frac{d}{dx}\frac{\varphi(y)-\varphi(x)}{y-x}\ge 0 \text{ or } \frac{d}{dy}\frac{\varphi(y)-\varphi(x)}{y-x}\ge 0 \text{ for } a<x<y<b.$$

For the first case,

$$\begin{aligned}\frac{d}{dx}\frac{\varphi(y)-\varphi(x)}{y-x}&=\frac{\varphi(y)-\varphi(x)-\varphi'(x)(y-x)}{(y-x)^2}\\&=\int_0^1\varphi''(x+t(y-x))(1-t)\,dt\ge 0.\end{aligned}$$

Similarly,

$$\frac{d}{dy}\frac{\varphi(y)-\varphi(x)}{y-x}=\frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2}$$

where we now use,

$$\varphi(x)-\varphi(y)=\varphi'(y)(x-y)+(x-y)^2\int_0^1\varphi''(y+t(x-y))(1-t)\,dt$$

so that

$$\frac{\varphi'(y)(y-x)-[\varphi(y)-\varphi(x)]}{(y-x)^2}=(x-y)^2\int_0^1\varphi''(y+t(x-y))(1-t)\,dt\ge 0$$

again.

3b. Let

$$f(t):=\varphi(u)+t(\varphi(v)-\varphi(u))-\varphi(u+t(v-u)).$$

Then $f(0)=f(1)=0$ with $\ddot f(t)=-(v-u)^2\varphi''(u+t(v-u))\le 0$ for almost all t. By the mean value theorem, there exists, $t_0\in(0,1)$ such that $\dot f(t_0)=0$ and then by the fundamental theorem of calculus it follows that

$$\dot f(t)=\int_{t_0}^t \ddot f(\tau)\,dt.$$

In particular, $\dot f(t)\le 0$ for $t>t_0$ and $\dot f(t)\ge 0$ for $t<t_0$ and hence $f(t)\ge f(1)=0$ for $t\ge t_0$ and $f(t)\ge f(0)=0$ for $t\le t_0$, i.e. $f(t)\ge 0$.

3c. Let $h:[0,1]\to\mathbb{R}$ be a piecewise C^2 – function. Then by the fundamental theorem of calculus and integration by parts,

$$h(t)=h(0)+\int_0^t h(\tau)\,d\tau=h(0)+th(t)-\int_0^t h(\tau)\,\tau d\tau$$

and

$$h(1)=h(t)+\int_t^1 h(\tau)\,d(\tau-1)=h(t)-(t-1)h(t)-\int_t^1 h(\tau)(\tau-1)\,d\tau.$$

Thus we have shown,

$$\begin{aligned}h(t)&=h(0)+th(t)-\int_0^t h(\tau)\,\tau d\tau \text{ and}\\h(t)&=h(1)+(t-1)h(t)+\int_t^1 h(\tau)(\tau-1)\,d\tau.\end{aligned}$$

So if we multiply the first equation by $(1-t)$ and add to it the second equation multiplied by t shows,

$$h(t)=(1-t)h(0)+th(1)-\int_0^1 G(t,\tau)\,\ddot h(\tau)\,d\tau, \tag{12.32}$$

where

$$G(t,\tau):=\begin{cases}\tau(1-t) \text{ if } \tau\le t\\ t(1-\tau) \text{ if } \tau\ge t\end{cases}.$$

(The function $G(t,\tau)$ is the "Green's function" for the operator $-d^2/dt^2$ on $[0,1]$ with Dirichlet boundary conditions. The formula in Eq. (12.32) is a standard representation formula for $h(t)$ which appears naturally in the study of harmonic functions.)

We now take $h(t):=\varphi(x_0+t(x_1-x_0))$ in Eq. (12.32) to learn

$$\varphi(x_0+t(x_1-x_0)) = (1-t)\varphi(x_0)+t\varphi(x_1) - (x_1-x_0)^2\int_0^1 G(t,\tau)\ddot{\varphi}(x_0+\tau(x_1-x_0))\,d\tau \le (1-t)\varphi(x_0)+t\varphi(x_1),$$

because $\ddot{\varphi}\ge 0$ and $G(t,\tau)\ge 0$. ■

Example 12.45. The functions $\exp(x)$ and $-\log(x)$ are convex and $|x|^p$ is convex iff $p\ge 1$ as follows from Lemma 12.44.

Example 12.46 (Proof of Lemma 10.36). Taking $\varphi(x)=e^{-x}$ in Lemma 12.44, Eq. (12.30) with $x_0=0$ implies (see Figure 10.1),

$$1-x\le\varphi(x)=e^{-x}\text{ for all }x\in\mathbb{R}.$$

Taking $\varphi(x)=e^{-2x}$ in Lemma 12.44, Eq. (12.28) with $x_0=0$ and $x_1=1$ implies, for all $t\in[0,1]$,

$$e^{-t}\le\varphi\left((1-t)0+t\frac{1}{2}\right) \le (1-t)\varphi(0)+t\varphi\left(\frac{1}{2}\right)=1-t+te^{-1}\le 1-\frac{1}{2}t,$$

wherein the last equality we used $e^{-1}<\frac{1}{2}$. Taking $t=2x$ in this equation then gives (see Figure 10.2)

$$e^{-2x}\le 1-x\text{ for }0\le x\le\frac{1}{2}. \tag{12.33}$$

Theorem 12.47. *Suppose that $\varphi:(a,b)\to\mathbb{R}$ is convex and for $x,y\in(a,b)$ with $x<y$, let*[8]

$$F(x,y):=\frac{\varphi(y)-\varphi(x)}{y-x}.$$

Then;

1. *$F(x,y)$ is increasing in each of its arguments.*
2. *The following limits exist,*

$$\varphi'_+(x):=F(x,x+):=\lim_{y\downarrow x}F(x,y)<\infty\text{ and} \tag{12.34}$$

$$\varphi'_-(y):=F(y-,y):=\lim_{x\uparrow y}F(x,y)>-\infty. \tag{12.35}$$

[8] The same formula would define $F(x,y)$ for $x\ne y$. However, since $F(x,y)=F(y,x)$, we would gain no new information by this extension.

3. *The functions, $\varphi'_\pm$ are both increasing functions and further satisfy,*

$$-\infty<\varphi'_-(x)\le\varphi'_+(x)\le\varphi'_-(y)<\infty\ \forall\ a<x<y<b. \tag{12.36}$$

4. *For any $t\in[\varphi'_-(x),\varphi'_+(x)]$,*

$$\varphi(y)\ge\varphi(x)+t(y-x)\text{ for all }x,y\in(a,b). \tag{12.37}$$

5. *For $a<\alpha<\beta<b$, let $K:=\max\{|\varphi'_+(\alpha)|,|\varphi'_-(\beta)|\}$. Then*

$$|\varphi(y)-\varphi(x)|\le K|y-x|\text{ for all }x,y\in[\alpha,\beta].$$

That is φ is Lipschitz continuous on $[\alpha,\beta]$.

6. *The function φ'_+ is right continuous and φ'_- is left continuous.*
7. *The set of discontinuity points for φ'_+ and for φ'_- are the same as the set of points of non-differentiability of φ. Moreover this set is at most countable.*

Proof. BRUCE: The first two items are a repetition of Lemma 12.42.

1. and 2. If we let $h_t=t\varphi(x_1)+(1-t)\varphi(x_0)$, then (x_t,h_t) is on the line segment joining $(x_0,\varphi(x_0))$ to $(x_1,\varphi(x_1))$ and the statement that φ is convex is then equivalent of $\varphi(x_t)\le h_t$ for all $0\le t\le 1$. Since

$$\frac{h_t-\varphi(x_0)}{x_t-x_0}=\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0}=\frac{\varphi(x_1)-h_t}{x_1-x_t},$$

the convexity of φ is equivalent to

$$\frac{\varphi(x_t)-\varphi(x_0)}{x_t-x_0}\le\frac{h_t-\varphi(x_0)}{x_t-x_0}=\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0}\text{ for all }x_0\le x_t\le x_1$$

and to

$$\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0}=\frac{\varphi(x_1)-h_t}{x_1-x_t}\le\frac{\varphi(x_1)-\varphi(x_t)}{x_1-x_t}\text{ for all }x_0\le x_t\le x_1.$$

Convexity also implies

$$\frac{\varphi(x_t)-\varphi(x_0)}{x_t-x_0}=\frac{h_t-\varphi(x_0)}{x_t-x_0}=\frac{\varphi(x_1)-h_t}{x_1-x_t}\le\frac{\varphi(x_1)-\varphi(x_t)}{x_1-x_t}.$$

These inequalities may be written more compactly as,

$$\frac{\varphi(v)-\varphi(u)}{v-u}\le\frac{\varphi(w)-\varphi(u)}{w-u}\le\frac{\varphi(w)-\varphi(v)}{w-v}, \tag{12.38}$$

valid for all $a<u<v<w<b$, again see Figure 12.2. The first (second) inequality in Eq. (12.38) shows $F(x,y)$ is increasing y (x). This then implies the limits in item 2. are monotone and hence exist as claimed.

3. Let $a < x < y < b$. Using the increasing nature of F,

$$-\infty < \varphi'_-(x) = F(x-, x) \le F(x, x+) = \varphi'_+(x) < \infty$$

and

$$\varphi'_+(x) = F(x, x+) \le F(y-, y) = \varphi'_-(y)$$

as desired.

4. Let $t \in \left[\varphi'_-(x), \varphi'_+(x)\right]$. Then

$$t \le \varphi'_+(x) = F(x, x+) \le F(x, y) = \frac{\varphi(y) - \varphi(x)}{y - x}$$

or equivalently,

$$\varphi(y) \ge \varphi(x) + t(y - x) \text{ for } y \ge x.$$

Therefore Eq. (12.37) holds for $y \ge x$. Similarly, for $y < x$,

$$t \ge \varphi'_-(x) = F(x-, x) \ge F(y, x) = \frac{\varphi(x) - \varphi(y)}{x - y}$$

or equivalently,

$$\varphi(y) \ge \varphi(x) - t(x - y) = \varphi(x) + t(y - x) \text{ for } y \le x.$$

Hence we have proved Eq. (12.37) for all $x, y \in (a, b)$.

5. For $a < \alpha \le x < y \le \beta < b$, we have

$$\varphi'_+(\alpha) \le \varphi'_+(x) = F(x, x+) \le F(x, y) \le F(y-, y) = \varphi'_-(y) \le \varphi'_-(\beta) \tag{12.39}$$

and in particular,

$$-K \le \varphi'_+(\alpha) \le \frac{\varphi(y) - \varphi(x)}{y - x} \le \varphi'_-(\beta) \le K.$$

This last inequality implies, $|\varphi(y) - \varphi(x)| \le K(y - x)$ which is the desired Lipschitz bound.

6. For $a < c < x < y < b$, we have $\varphi'_+(x) = F(x, x+) \le F(x, y)$ and letting $x \downarrow c$ (using the continuity of F) we learn $\varphi'_+(c+) \le F(c, y)$. We may now let $y \downarrow c$ to conclude $\varphi'_+(c+) \le \varphi'_+(c)$. Since $\varphi'_+(c) \le \varphi'_+(c+)$, it follows that $\varphi'_+(c) = \varphi'_+(c+)$ and hence that φ'_+ is right continuous.

Similarly, for $a < x < y < c < b$, we have $\varphi'_-(y) \ge F(x, y)$ and letting $y \uparrow c$ (using the continuity of F) we learn $\varphi'_-(c-) \ge F(x, c)$. Now let $x \uparrow c$ to conclude $\varphi'_-(c-) \ge \varphi'_-(c)$. Since $\varphi'_-(c) \ge \varphi'_-(c-)$, it follows that $\varphi'_-(c) = \varphi'_-(c-)$, i.e. φ'_- is left continuous.

7. Since $\varphi_\pm$ are increasing functions, they have at most countably many points of discontinuity. Letting $x \uparrow y$ in Eq. (12.36), using the left continuity of φ'_-, shows $\varphi'_-(y) = \varphi'_+(y-)$. Hence if φ'_- is continuous at y, $\varphi'_-(y) = \varphi'_-(y+) = \varphi'_+(y)$ and φ is differentiable at y. Conversely if φ is differentiable at y, then

$$\varphi'_+(y-) = \varphi'_-(y) = \varphi'(y) = \varphi'_+(y)$$

which shows φ'_+ is continuous at y. Thus we have shown that set of discontinuity points of φ'_+ is the same as the set of points of non-differentiability of φ. That the discontinuity set of φ'_- is the same as the non-differentiability set of φ is proved similarly. ■

Corollary 12.48. *If $\varphi : (a, b) \to \mathbb{R}$ is a convex function and $D \subset (a, b)$ is a dense set, then*

$$\varphi(y) = \sup_{x \in D}\left[\varphi(x) + \varphi'_\pm(x)(y - x)\right] \text{ for all } x, y \in (a, b).$$

Proof. Let $\psi_\pm(y) := \sup_{x \in D}\left[\varphi(x) + \varphi_\pm(x)(y - x)\right]$. According to Eq. (12.37) above, we know that $\varphi(y) \ge \psi_\pm(y)$ for all $y \in (a, b)$. Now suppose that $x \in (a, b)$ and $x_n \in \Lambda$ with $x_n \uparrow x$. Then passing to the limit in the estimate, $\psi_-(y) \ge \varphi(x_n) + \varphi'_-(x_n)(y - x_n)$, shows $\psi_-(y) \ge \varphi(x) + \varphi'_-(x)(y - x)$. Since $x \in (a, b)$ is arbitrary we may take $x = y$ to discover $\psi_-(y) \ge \varphi(y)$ and hence $\varphi(y) = \psi_-(y)$. The proof that $\varphi(y) = \psi_+(y)$ is similar. ■

Lemma 12.49. *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is a non-decreasing function such that*

$$\varphi\left(\frac{1}{2}(x + y)\right) \le \frac{1}{2}\left[\varphi(x) + \varphi(y)\right] \text{ for all } x, y \in (a, b), \tag{12.40}$$

then φ is convex. The result remains true if φ is assumed to be continuous rather than non-decreasing.

Proof. Let $x_0, x_1 \in (a, b)$ and $x_t := x_0 + t(x_1 - x_0)$ as above. For $n \in \mathbb{N}$ let $\mathbb{D}_n = \left\{\frac{k}{2^n} : 1 \le k < 2^n\right\}$. We are going to being by showing Eq. (12.40) implies

$$\varphi(x_t) \le (1 - t)\varphi(x_0) + t\varphi(x_1) \text{ for all } t \in \mathbb{D} := \cup_n \mathbb{D}_n. \tag{12.41}$$

We will do this by induction on n. For $n = 1$, this follows directly from Eq. (12.40). So now suppose that Eq. (12.41) holds for all $t \in \mathbb{D}_n$ and now suppose that $t = \frac{2k+1}{2^n} \in \mathbb{D}_{n+1}$. Observing that

$$x_t = \frac{1}{2}\left(x_{\frac{k}{2^{n-1}}} + x_{\frac{k+1}{2^n}}\right)$$

we may again use Eq. (12.40) to conclude,

$$\varphi(x_t) \le \frac{1}{2}\left(\varphi\left(x_{\frac{k}{2^{n-1}}}\right) + \varphi\left(x_{\frac{k+1}{2^{n-1}}}\right)\right).$$

Then use the induction hypothesis to conclude,

$$\begin{aligned}\varphi\left(x_t\right) &\leq \frac{1}{2}\begin{pmatrix}\left(1-\frac{k}{2^{n-1}}\right)\varphi\left(x_0\right)+\frac{k}{2^{n-1}}\varphi\left(x_1\right)\\+\left(1-\frac{k+1}{2^{n-1}}\right)\varphi\left(x_0\right)+\frac{k+1}{2^{n-1}}\varphi\left(x_1\right)\end{pmatrix}\\&=\left(1-t\right)\varphi\left(x_0\right)+t\varphi\left(x_1\right)\end{aligned}$$

as desired.

For general $t \in (0,1)$, let $\tau \in \mathbb{D}$ such that $\tau > t$. Since φ is increasing and by Eq. (12.41) we conclude,

$$\varphi\left(x_t\right) \leq \varphi\left(x_\tau\right) \leq \left(1-\tau\right)\varphi\left(x_0\right)+\tau\varphi\left(x_1\right).$$

We may now let $\tau \downarrow t$ to complete the proof. This same technique clearly also works if we were to assume that φ is continuous rather than monotonic. ■

13

Hilbert Space Basics

Definition 13.1. *Let H be a complex vector space. An inner product on H is a function, $\langle\cdot|\cdot\rangle : H \times H \to \mathbb{C}$, such that*

1. $\langle ax + by|z\rangle = a\langle x|z\rangle + b\langle y|z\rangle$ *i.e.* $x \to \langle x|z\rangle$ *is linear.*
2. $\overline{\langle x|y\rangle} = \langle y|x\rangle$.
3. $\|x\|^2 := \langle x|x\rangle \geq 0$ *with equality* $\|x\|^2 = 0$ *iff* $x = 0$.

Notice that combining properties (1) and (2) that $x \to \langle z|x\rangle$ is conjugate linear for fixed $z \in H$, i.e.

$$\langle z|ax + by\rangle = \bar{a}\langle z|x\rangle + \bar{b}\langle z|y\rangle.$$

The following identity will be used frequently in the sequel without further mention,

$$\begin{aligned} \|x + y\|^2 &= \langle x + y|x + y\rangle = \|x\|^2 + \|y\|^2 + \langle x|y\rangle + \langle y|x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle. \end{aligned} \tag{13.1}$$

Theorem 13.2 (Schwarz Inequality). *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, then for all $x, y \in H$*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

and equality holds iff x and y are linearly dependent.

Proof. If $y = 0$, the result holds trivially. So assume that $y \neq 0$ and observe; if $x = \alpha y$ for some $\alpha \in \mathbb{C}$, then $\langle x|y\rangle = \bar{\alpha}\,\|y\|^2$ and hence

$$|\langle x|y\rangle| = |\alpha|\,\|y\|^2 = \|x\|\|y\|.$$

Now suppose that $x \in H$ is arbitrary, let $z := x - \|y\|^{-2}\langle x|y\rangle y$. (So $\|y\|^{-2}\langle x|y\rangle y$ is the "orthogonal projection" of x along y, see Figure 13.1.) Then

$$\begin{aligned} 0 \leq \|z\|^2 &= \left\| x - \frac{\langle x|y\rangle}{\|y\|^2} y \right\|^2 = \|x\|^2 + \frac{|\langle x|y\rangle|^2}{\|y\|^4}\|y\|^2 - 2\mathrm{Re}\langle x|\frac{\langle x|y\rangle}{\|y\|^2} y\rangle \\ &= \|x\|^2 - \frac{|\langle x|y\rangle|^2}{\|y\|^2} \end{aligned}$$

from which it follows that $0 \leq \|y\|^2\|x\|^2 - |\langle x|y\rangle|^2$ with equality iff $z = 0$ or equivalently iff $x = \|y\|^{-2}\langle x|y\rangle y$. ■

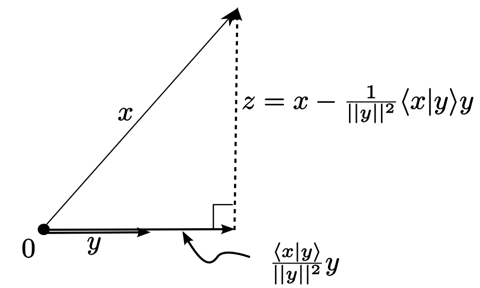


Fig. 13.1. The picture behind the proof of the Schwarz inequality.

Corollary 13.3. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space and $\|x\| := \sqrt{\langle x|x\rangle}$. Then the **Hilbertian norm**, $\|\cdot\|$, is a norm on H. Moreover $\langle\cdot|\cdot\rangle$ is continuous on $H \times H$, where H is viewed as the normed space $(H, \|\cdot\|)$.*

Proof. If $x, y \in H$, then, using Schwarz's inequality,

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2. \end{aligned}$$

Taking the square root of this inequality shows $\|\cdot\|$ satisfies the triangle inequality.

Checking that $\|\cdot\|$ satisfies the remaining axioms of a norm is now routine and will be left to the reader. If $x, y, \Delta x, \Delta y \in H$, then

$$\begin{aligned} |\langle x + \Delta x|y + \Delta y\rangle - \langle x|y\rangle| &= |\langle x|\Delta y\rangle + \langle \Delta x|y\rangle + \langle \Delta x|\Delta y\rangle| \\ &\leq \|x\|\|\Delta y\| + \|y\|\|\Delta x\| + \|\Delta x\|\|\Delta y\| \\ &\to 0 \text{ as } \Delta x, \Delta y \to 0, \end{aligned}$$

from which it follows that $\langle\cdot|\cdot\rangle$ is continuous. ■

Definition 13.4. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space, we say $x, y \in H$ are **orthogonal** and write $x \perp y$ iff $\langle x|y\rangle = 0$. More generally if $A \subset H$ is a set, $x \in H$ is **orthogonal to** A (write $x \perp A$) iff $\langle x|y\rangle = 0$ for all $y \in A$. Let $A^\perp = \{x \in H : x \perp A\}$ be the set of vectors orthogonal to A. A subset $S \subset H$*

is an ***orthogonal set*** *if $x \perp y$ for all distinct elements $x, y \in S$. If S further satisfies, $\|x\| = 1$ for all $x \in S$, then S is said to be an* ***orthonormal set.***

Proposition 13.5. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. ***(Parallelogram Law)***
$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{13.2}$$
for all $x, y \in H$.
2. ***(Pythagorean Theorem)*** *If $S \subset\subset H$ is a finite orthogonal set, then*
$$\left\|\sum_{x \in S} x\right\|^2 = \sum_{x \in S} \|x\|^2. \tag{13.3}$$
3. *If $A \subset H$ is a set, then $A^\perp$ is a* ***closed*** *linear subspace of H.*

Proof. I will assume that H is a complex Hilbert space, the real case being easier. Items 1. and 2. are proved by the following elementary computations;

$$\begin{aligned}
&\|x+y\|^2 + \|x-y\|^2 \\
&\quad = \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle + \|x\|^2 + \|y\|^2 - 2\mathrm{Re}\langle x|y\rangle \\
&\quad = 2\|x\|^2 + 2\|y\|^2,
\end{aligned}$$

and

$$\begin{aligned}
\left\|\sum_{x \in S} x\right\|^2 &= \langle \sum_{x \in S} x | \sum_{y \in S} y\rangle = \sum_{x,y \in S} \langle x|y\rangle \\
&= \sum_{x \in S} \langle x|x\rangle = \sum_{x \in S} \|x\|^2.
\end{aligned}$$

Item 3. is a consequence of the continuity of $\langle\cdot|\cdot\rangle$ and the fact that

$$A^\perp = \cap_{x \in A} \mathrm{Nul}(\langle\cdot|x\rangle)$$

where $\mathrm{Nul}(\langle\cdot|x\rangle) = \{y \in H : \langle y|x\rangle = 0\}$ – a closed subspace of H. ■

Definition 13.6. *A* ***Hilbert space*** *is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

Example 13.7. For any measure space, $(\Omega, \mathcal{B}, \mu)$, $H := L^2(\mu)$ with inner product,

$$\langle f|g\rangle = \int_\Omega f(\omega)\,\bar{g}(\omega)\,d\mu(\omega)$$

is a Hilbert space – see Theorem 12.17 for the completeness assertion.

Definition 13.8. *A subset C of a vector space X is said to be convex if for all $x, y \in C$ the line segment $[x, y] := \{tx + (1-t)y : 0 \le t \le 1\}$ joining x to y is contained in C as well. (Notice that any vector subspace of X is convex.)*

Theorem 13.9 (Best Approximation Theorem). *Suppose that H is a Hilbert space and $M \subset H$ is a closed convex subset of H. Then for any $x \in H$ there exists a unique $y \in M$ such that*

$$\|x-y\| = d(x, M) = \inf_{z \in M} \|x-z\|.$$

Moreover, if M is a vector subspace of H, then the point y may also be characterized as the unique point in M such that $(x-y) \perp M$.

Proof. Let $x \in H$, $\delta := d(x, M)$, $y, z \in M$, and, referring to Figure 13.2, let $w = z + (y - x)$ and $c = (x+y)/2 \in M$. It then follows by the parallelogram law (Eq. (13.2) with $x \to (y-x)$ and $y \to (z-x)$) and the fact that $c \in M$ that

$$\begin{aligned}
2\|y-x\|^2 + 2\|z-x\|^2 &= \|w-x\|^2 + \|y-z\|^2 \\
&= \|z+y-2x\|^2 + \|y-z\|^2 \\
&= 4\|x-c\|^2 + \|y-z\|^2 \\
&\ge 4\delta^2 + \|y-z\|^2.
\end{aligned}$$

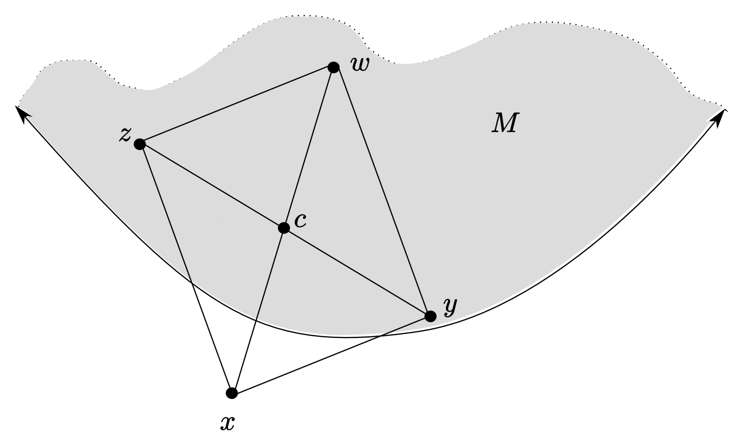


Fig. 13.2. In this figure $y, z \in M$ and by convexity, $c = (x+y)/2 \in M$.

Thus we have shown for all $y, z \in M$ that,

$$\|y-z\|^2 \le 2\|y-x\|^2 + 2\|z-x\|^2 - 4\delta^2. \tag{13.4}$$

Uniqueness. If $y, z \in M$ minimize the distance to x, then $\|y-x\| = \delta = \|z-x\|$ and it follows from Eq. (13.4) that $y = z$.

Existence. Let $y_n \in M$ be chosen such that $\|y_n - x\| = \delta_n \to \delta = d(x, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (13.4) shows

$$\|y_n - y_m\|^2 \le 2\delta_m^2 + 2\delta_n^2 - 4\delta^2 \to 0 \text{ as } m, n \to \infty.$$

Therefore, by completeness of H, $\{y_n\}_{n=1}^{\infty}$ is convergent. Because M is closed, $y := \lim_{n\to\infty} y_n \in M$ and because the norm is continuous,

$$\|y - x\| = \lim_{n\to\infty} \|y_n - x\| = \delta = d(x, M).$$

So y is the desired point in M which is closest to x.

Orthogonality property. Now suppose M is a closed subspace of H and $x \in H$. Let $y \in M$ be the closest point in M to x. Then for $w \in M$, the function

$$g(t) := \|x - (y + tw)\|^2 = \|x - y\|^2 - 2t\mathrm{Re}\langle x - y|w\rangle + t^2\|w\|^2$$

has a minimum at $t = 0$ and therefore $0 = g'(0) = -2\mathrm{Re}\langle x - y|w\rangle$. Since $w \in M$ is arbitrary, this implies that $(x - y) \perp M$, see Figure 13.3.

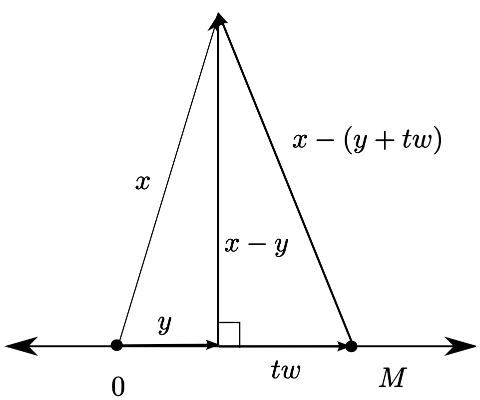


Fig. 13.3. The orthogonality relationships of closest points.

Finally suppose $y \in M$ is any point such that $(x - y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x - z\|^2 = \|x - y + y - z\|^2 = \|x - y\|^2 + \|y - z\|^2 \ge \|x - y\|^2$$

which shows $d(x, M)^2 \ge \|x - y\|^2$. That is to say y is the point in M closest to x. ■

Definition 13.10. *Suppose that $A : H \to H$ is a bounded operator, i.e.*

$$\|A\| := \sup\{\|Ax\| : x \in H \text{ with } \|x\| = 1\} < \infty.$$

*The **adjoint** of A, denoted A^*, is the unique operator $A^* : H \to H$ such that $\langle Ax|y\rangle = \langle x|A^*y\rangle$. (The proof that A^* exists and is unique will be given in Proposition 13.15 below.) A bounded operator $A : H \to H$ is **self - adjoint** or **Hermitian** if $A = A^*$.*

Definition 13.11. *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection of H onto M is the function $P_M : H \to H$ such that for $x \in H$, $P_M(x)$ is the unique element in M such that $(x - P_M(x)) \perp M$, i.e. $P_M(x)$ is the unique element in M such that*

$$\langle x|m\rangle = \langle P_M(x)|m\rangle \text{ for all } m \in M. \tag{13.5}$$

Given a linear transformation A, we will let $\mathrm{Ran}\,(A)$ and $\mathrm{Nul}\,(A)$ denote the **range** and the **null-space** of A respectively.

Theorem 13.12 (Projection Theorem). *Let H be a Hilbert space and $M \subset H$ be a closed subspace. The orthogonal projection P_M satisfies:*

1. *P_M is linear and hence we will write $P_M x$ rather than $P_M(x)$.*
2. *$P_M^2 = P_M$ (P_M is a projection).*
3. *$P_M^* = P_M$ (P_M is self-adjoint).*
4. *$\mathrm{Ran}(P_M) = M$ and $\mathrm{Nul}(P_M) = M^\perp$.*
5. *If $N \subset M \subset H$ is another closed subspace, the $P_N P_M = P_M P_N = P_N$.*

Proof.

1. Let $x_1, x_2 \in H$ and $\alpha \in \mathbb{C}$, then $P_M x_1 + \alpha P_M x_2 \in M$ and
$$P_M x_1 + \alpha P_M x_2 - (x_1 + \alpha x_2) = [P_M x_1 - x_1 + \alpha(P_M x_2 - x_2)] \in M^\perp$$
showing $P_M x_1 + \alpha P_M x_2 = P_M(x_1 + \alpha x_2)$, i.e. P_M is linear.
2. Obviously $\mathrm{Ran}(P_M) = M$ and $P_M x = x$ for all $x \in M$. Therefore $P_M^2 = P_M$.
3. Let $x, y \in H$, then since $(x - P_M x)$ and $(y - P_M y)$ are in $M^\perp$,
$$\begin{aligned}\langle P_M x|y\rangle &= \langle P_M x|P_M y + y - P_M y\rangle = \langle P_M x|P_M y\rangle \\ &= \langle P_M x + (x - P_M x)|P_M y\rangle = \langle x|P_M y\rangle.\end{aligned}$$
4. We have already seen, $\mathrm{Ran}(P_M) = M$ and $P_M x = 0$ iff $x = x - 0 \in M^\perp$, i.e. $\mathrm{Nul}(P_M) = M^\perp$.

5. If $N \subset M \subset H$ it is clear that $P_M P_N = P_N$ since $P_M = Id$ on $N = \mathrm{Ran}(P_N) \subset M$. Taking adjoints gives the other identity, namely that $P_N P_M = P_N$. More directly, if $x \in H$ and $n \in N$, we have

$$\langle P_N P_M x | n \rangle = \langle P_M x | P_N n \rangle = \langle P_M x | n \rangle = \langle x | P_M n \rangle = \langle x | n \rangle .$$

Since this holds for all n we may conclude that $P_N P_M x = P_N x$.

■

Corollary 13.13. *If $M \subset H$ is a proper closed subspace of a Hilbert space H, then $H = M \oplus M^\perp$.*

Proof. Given $x \in H$, let $y = P_M x$ so that $x - y \in M^\perp$. Then $x = y + (x-y) \in M + M^\perp$. If $x \in M \cap M^\perp$, then $x \perp x$, i.e. $\|x\|^2 = \langle x|x\rangle = 0$. So $M \cap M^\perp = \{0\}$. ■

Exercise 13.1. Suppose M is a subset of H, then $M^{\perp\perp} = \overline{\mathrm{span}(M)}$.

Theorem 13.14 (Riesz Theorem). *Let H^* be the dual space of H (i.e. that linear space of continuous linear functionals on H). The map*

$$z \in H \xrightarrow{j} \langle \cdot | z \rangle \in H^* \tag{13.6}$$

is a conjugate linear[1] *isometric isomorphism.*

Proof. The map j is conjugate linear by the axioms of the inner products. Moreover, for $x, z \in H$,

$$|\langle x|z\rangle| \le \|x\| \, \|z\| \text{ for all } x \in H$$

with equality when $x = z$. This implies that $\|jz\|_{H^*} = \|\langle \cdot|z\rangle\|_{H^*} = \|z\|$. Therefore j is isometric and this implies j is injective. To finish the proof we must show that j is surjective. So let $f \in H^*$ which we assume, without loss of generality, is non-zero. Then $M = \mathrm{Nul}(f)$ – a closed proper subspace of H. Since, by Corollary 13.13, $H = M \oplus M^\perp$, $f : H/M \cong M^\perp \to \mathbb{F}$ is a linear isomorphism. This shows that $\dim(M^\perp) = 1$ and hence $H = M \oplus \mathbb{F} x_0$ where $x_0 \in M^\perp \setminus \{0\}$.[2] Choose $z = \lambda x_0 \in M^\perp$ such that $f(x_0) = \langle x_0 | z \rangle$, i.e. $\lambda = \bar{f}(x_0)/\|x_0\|^2$. Then for $x = m + \lambda x_0$ with $m \in M$ and $\lambda \in \mathbb{F}$,

$$f(x) = \lambda f(x_0) = \lambda \langle x_0|z\rangle = \langle \lambda x_0 | z\rangle = \langle m + \lambda x_0 | z\rangle = \langle x|z\rangle$$

which shows that $f = jz$. ■

[1] Recall that j is conjugate linear if

$$j(z_1 + \alpha z_2) = jz_1 + \bar{\alpha} j z_2$$

for all $z_1, z_2 \in H$ and $\alpha \in \mathbb{C}$.

[2] Alternatively, choose $x_0 \in M^\perp \setminus \{0\}$ such that $f(x_0) = 1$. For $x \in M^\perp$ we have $f(x - \lambda x_0) = 0$ provided that $\lambda := f(x)$. Therefore $x - \lambda x_0 \in M \cap M^\perp = \{0\}$, i.e. $x = \lambda x_0$. This again shows that $M^\perp$ is spanned by x_0.

Proposition 13.15 (Adjoints). *Let H and K be Hilbert spaces and $A : H \to K$ be a bounded operator. Then there exists a unique bounded operator $A^* : K \to H$ such that*

$$\langle Ax|y\rangle_K = \langle x|A^* y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{13.7}$$

Moreover, for all $A, B \in L(H,K)$ and $\lambda \in \mathbb{C}$,

1. $(A + \lambda B)^* = A^* + \bar{\lambda} B^*$,
2. $A^{**} := (A^*)^* = A$,
3. $\|A^*\| = \|A\|$ *and*
4. $\|A^* A\| = \|A\|^2$.
5. *If $K = H$, then $(AB)^* = B^* A^*$. In particular $A \in L(H)$ has a bounded inverse iff A^* has a bounded inverse and $(A^*)^{-1} = (A^{-1})^*$.*

Proof. For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in H^* and therefore there exists, by Theorem 13.14, a unique vector $z \in H$ (we will denote this z by $A^*(y)$) such that

$$\langle Ax|y\rangle_K = \langle x|z\rangle_H \text{ for all } x \in H.$$

This shows there is a unique map $A^* : K \to H$ such that $\langle Ax|y\rangle_K = \langle x|A^*(y)\rangle_H$ for all $x \in H$ and $y \in K$.

To see A^* is linear, let $y_1, y_2 \in K$ and $\lambda \in \mathbb{C}$, then for any $x \in H$,

$$\begin{aligned}
\langle Ax|y_1 + \lambda y_2\rangle_K &= \langle Ax|y_1\rangle_K + \bar{\lambda}\langle Ax|y_2\rangle_K \\
&= \langle x|A^*(y_1)\rangle_K + \bar{\lambda}\langle x|A^*(y_2)\rangle_H \\
&= \langle x|A^*(y_1) + \lambda A^*(y_2)\rangle_H
\end{aligned}$$

and by the uniqueness of $A^*(y_1 + \lambda y_2)$ we find

$$A^*(y_1 + \lambda y_2) = A^*(y_1) + \lambda A^*(y_2).$$

This shows A^* is linear and so we will now write $A^* y$ instead of $A^*(y)$.

Since

$$\langle A^* y | x\rangle_H = \overline{\langle x|A^* y\rangle_H} = \overline{\langle Ax|y\rangle_K} = \langle y|Ax\rangle_K$$

it follows that $A^{**} = A$. The assertion that $(A + \lambda B)^* = A^* + \bar{\lambda} B^*$ is Exercise 13.2.

Items 3. and 4. Making use of Schwarz's inequality (Theorem 13.2), we have

$$\begin{aligned}\|A^*\| &= \sup_{k\in K:\|k\|=1} \|A^*k\| \\ &= \sup_{k\in K:\|k\|=1}\ \sup_{h\in H:\|h\|=1} |\langle A^*k|h\rangle| \\ &= \sup_{h\in H:\|h\|=1}\ \sup_{k\in K:\|k\|=1} |\langle k|Ah\rangle| = \sup_{h\in H:\|h\|=1} \|Ah\| = \|A\|\end{aligned}$$

so that $\|A^*\| = \|A\|$. Since

$$\|A^*A\| \le \|A^*\|\,\|A\| = \|A\|^2$$

and

$$\begin{aligned}\|A\|^2 &= \sup_{h\in H:\|h\|=1} \|Ah\|^2 = \sup_{h\in H:\|h\|=1} |\langle Ah|Ah\rangle| \\ &= \sup_{h\in H:\|h\|=1} |\langle h|A^*Ah\rangle| \le \sup_{h\in H:\|h\|=1} \|A^*Ah\| = \|A^*A\| \qquad (13.8)\end{aligned}$$

we also have $\|A^*A\| \le \|A\|^2 \le \|A^*A\|$ which shows $\|A\|^2 = \|A^*A\|$.

Alternatively, from Eq. (13.8),

$$\|A\|^2 \le \|A^*A\| \le \|A\|\,\|A^*\| \qquad (13.9)$$

which then implies $\|A\| \le \|A^*\|$. Replacing A by A^* in this last inequality shows $\|A^*\| \le \|A\|$ and hence that $\|A^*\| = \|A\|$. Using this identity back in Eq. (13.9) proves $\|A\|^2 = \|A^*A\|$.

Now suppose that $K = H$. Then

$$\langle ABh|k\rangle = \langle Bh|A^*k\rangle = \langle h|B^*A^*k\rangle$$

which shows $(AB)^* = B^*A^*$. If A^{-1} exists then

$$\begin{aligned}\left(A^{-1}\right)^* A^* &= \left(AA^{-1}\right)^* = I^* = I \text{ and} \\ A^*\left(A^{-1}\right)^* &= \left(A^{-1}A\right)^* = I^* = I.\end{aligned}$$

This shows that A^* is invertible and $(A^*)^{-1} = \left(A^{-1}\right)^*$. Similarly if A^* is invertible then so is $A = A^{**}$. ■

Exercise 13.2. Let H, K, M be Hilbert spaces, $A, B \in L(H,K)$, $C \in L(K,M)$ and $\lambda \in \mathbb{C}$. Show $(A+\lambda B)^* = A^* + \bar{\lambda}B^*$ and $(CA)^* = A^*C^* \in L(M,H)$.

Exercise 13.3. Let $H = \mathbb{C}^n$ and $K = \mathbb{C}^m$ equipped with the usual inner products, i.e. $\langle z|w\rangle_H = z\cdot\bar{w}$ for $z, w \in H$. Let A be an $m\times n$ matrix thought of as a linear operator from H to K. Show the matrix associated to $A^* : K \to H$ is the conjugate transpose of A.

Lemma 13.16. *Suppose $A : H \to K$ is a bounded operator, then:*

1. $\mathrm{Nul}(A^*) = \mathrm{Ran}(A)^\perp$.
2. $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$.
3. *if $K = H$ and $V \subset H$ is an A – invariant subspace (i.e. $A(V) \subset V$), then $V^\perp$ is A^* – invariant.*

Proof. An element $y \in K$ is in $\mathrm{Nul}(A^*)$ iff $0 = \langle A^*y|x\rangle = \langle y|Ax\rangle$ for all $x \in H$ which happens iff $y \in \mathrm{Ran}(A)^\perp$. Because, by Exercise 13.1, $\overline{\mathrm{Ran}(A)} = \mathrm{Ran}(A)^{\perp\perp}$, and so by the first item, $\overline{\mathrm{Ran}(A)} = \mathrm{Nul}(A^*)^\perp$. Now suppose $A(V) \subset V$ and $y \in V^\perp$, then

$$\langle A^*y|x\rangle = \langle y|Ax\rangle = 0 \text{ for all } x \in V$$

which shows $A^*y \in V^\perp$. ■

The next elementary theorem (referred to as the bounded linear transformation theorem, or B.L.T. theorem for short) is often useful.

Theorem 13.17 (B. L. T. Theorem). *Suppose that Z is a normed space, X is a Banach[3] space, and $\mathcal{S} \subset Z$ is a dense linear subspace of Z. If $T : \mathcal{S} \to X$ is a bounded linear transformation (i.e. there exists $C < \infty$ such that $\|Tz\| \le C\,\|z\|$ for all $z \in \mathcal{S}$), then T has a unique extension to an element $\bar{T} \in L(Z,X)$ and this extension still satisfies*

$$\left\|\bar{T}z\right\| \le C\,\|z\| \text{ for all } z \in \bar{\mathcal{S}}.$$

Proof. Let $z \in Z$ and choose $z_n \in \mathcal{S}$ such that $z_n \to z$. Since

$$\|Tz_m - Tz_n\| \le C\,\|z_m - z_n\| \to 0 \text{ as } m,n \to \infty,$$

it follows by the completeness of X that $\lim_{n\to\infty} Tz_n =: \bar{T}z$ exists. Moreover, if $w_n \in \mathcal{S}$ is another sequence converging to z, then

$$\|Tz_n - Tw_n\| \le C\,\|z_n - w_n\| \to C\,\|z - z\| = 0$$

and therefore $\bar{T}z$ is well defined. It is now a simple matter to check that $\bar{T} : Z \to X$ is still linear and that

$$\left\|\bar{T}z\right\| = \lim_{n\to\infty} \|Tz_n\| \le \lim_{n\to\infty} C\,\|z_n\| = C\,\|z\| \text{ for all } x \in Z.$$

Thus $\bar{T}$ is an extension of T to all of the Z. The uniqueness of this extension is easy to prove and will be left to the reader. ■

[3] A Banach space is a complete normed space. The main examples for us are Hilbert spaces.

13.1 Compactness Results for L^p – Spaces*

In this section we are going to identify the sequentially "weak" compact subsets of $L^p(\Omega,\mathcal{B},P)$ for $1\le p<\infty$, where $(\Omega,\mathcal{B},P)$ is a probability space. The key to our proofs will be the following Hilbert space compactness result.

Theorem 13.18. *Suppose $\{x_n\}_{n=1}^\infty$ is a bounded sequence in H (i.e. $C := \sup_n \|x_n\| < \infty$), then there exists a sub-sequence, $y_k := x_{n_k}$ and an $x\in H$ such that $\lim_{k\to\infty}\langle y_k|h\rangle = \langle x|h\rangle$ for all $h\in H$. We say that y_k converges to x weakly in this case and denote this by $y_k \overset{w}{\to} x$.*

Proof. Let $H_0 := \overline{\operatorname{span}(x_k : k\in\mathbb{N})}$. Then H_0 is a closed separable Hilbert subspace of H and $\{x_k\}_{k=1}^\infty \subset H_0$. Let $\{h_n\}_{n=1}^\infty$ be a countable dense subset of H_0. Since $|\langle x_k|h_n\rangle| \le \|x_k\|\,\|h_n\| \le C\,\|h_n\| < \infty$, the sequence, $\{\langle x_k|h_n\rangle\}_{k=1}^\infty \subset \mathbb{C}$, is bounded and hence has a convergent sub-sequence for all $n\in\mathbb{N}$. By the Cantor's diagonalization argument we can find a a sub-sequence, $y_k := x_{n_k}$, of $\{x_n\}$ such that $\lim_{k\to\infty}\langle y_k|h_n\rangle$ exists for all $n\in\mathbb{N}$.

We now show $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists for all $z\in H_0$. Indeed, for any $k,l,n\in\mathbb{N}$, we have

$$\begin{aligned}|\langle y_k|z\rangle - \langle y_l|z\rangle| &= |\langle y_k - y_l|z\rangle| \le |\langle y_k-y_l|h_n\rangle| + |\langle y_k - y_l|z-h_n\rangle| \\ &\le |\langle y_k - y_l|h_n\rangle| + 2C\,\|z-h_n\|\,.\end{aligned}$$

Letting $k,l\to\infty$ in this estimate then shows

$$\limsup_{k,l\to\infty} |\langle y_k|z\rangle - \langle y_l|z\rangle| \le 2C\,\|z-h_n\|\,.$$

Since we may choose $n\in\mathbb{N}$ such that $\|z-h_n\|$ is as small as we please, we may conclude that $\limsup_{k,l\to\infty}|\langle y_k|z\rangle - \langle y_l|z\rangle|$, i.e. $\varphi(z) := \lim_{k\to\infty}\langle y_k|z\rangle$ exists.

The function, $\bar\varphi(z) = \lim_{k\to\infty}\langle z|y_k\rangle$ is a bounded linear functional on H because

$$|\bar\varphi(z)| = \liminf_{k\to\infty}|\langle z|y_k\rangle| \le C\,\|z\|\,.$$

Therefore by the Riesz Theorem 13.14, there exists $x\in H_0$ such that $\bar\varphi(z) = \langle z|x\rangle$ for all $z\in H_0$. Thus, for this $x\in H_0$ we have shown

$$\lim_{k\to\infty}\langle y_k|z\rangle = \langle x|z\rangle \text{ for all } z\in H_0. \tag{13.10}$$

To finish the proof we need only observe that Eq. (13.10) is valid for all $z\in H$. Indeed if $z\in H$, then $z = z_0 + z_1$ where $z_0 = P_{H_0}z\in H_0$ and $z_1 = z - P_{H_0}z \in H_0^\perp$. Since $y_k,\ x\in H_0$, we have

$$\lim_{k\to\infty}\langle y_k|z\rangle = \lim_{k\to\infty}\langle y_k|z_0\rangle = \langle x|z_0\rangle = \langle x|z\rangle \text{ for all } z\in H.$$

■

Since unbounded subsets of H are clearly not sequentially weakly compact, Theorem 13.18 states that a set is sequentially precompact in H iff it is bounded. Let us now use Theorem 13.18 to identify the sequentially compact subsets of $L^p(\Omega,\mathcal{B},P)$ for all $1\le p<\infty$. We begin with the case $p=1$.

Theorem 13.19. *If $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega,\mathcal{B},P)$, there exists a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ and $X\in L^1(\Omega,\mathcal{B},P)$ such that*

$$\lim_{k\to\infty}\mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h\in\mathcal{B}_b. \tag{13.11}$$

Proof. For each $m\in\mathbb{N}$ let $X_n^m := X_n 1_{|X_n|\le m}$. The truncated sequence $\{X_n^m\}_{n=1}^\infty$ is a bounded subset of the Hilbert space, $L^2(\Omega,\mathcal{B},P)$, for all $m\in\mathbb{N}$. Therefore by Theorem 13.18, $\{X_n^m\}_{n=1}^\infty$ has a weakly convergent sub-sequence for all $m\in\mathbb{N}$. By Cantor's diagonalization argument, we can find $Y_k^m := X_{n_k}^m$ and $X^m\in L^2(\Omega,\mathcal{B},P)$ such that $Y_k^m \overset{w}{\to} X^m$ as $m\to\infty$ and in particular

$$\lim_{k\to\infty}\mathbb{E}[Y_k^m h] = \mathbb{E}[X^m h] \text{ for all } h\in\mathcal{B}_b.$$

Our next goal is to show $X^m \to X$ in $L^1(\Omega,\mathcal{B},P)$. To this end, for $m<M$ and $h\in\mathcal{B}_b$ we have

$$\begin{aligned}\left|\mathbb{E}\left[\left(X^M - X^m\right)h\right]\right| &= \lim_{k\to\infty}\left|\mathbb{E}\left[\left(Y_k^M - Y_k^m\right)h\right]\right| \le \liminf_{k\to\infty}\mathbb{E}\left[\left|Y_k^M - Y_k^m\right||h|\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : M\ge |Y_k| > m\right] \\ &\le \|h\|_\infty \cdot \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right].\end{aligned}$$

Taking $h = \overline{\operatorname{sgn}(X^M - X^m)}$ in this inequality shows

$$\mathbb{E}\left[\left|X^M - X^m\right|\right] \le \liminf_{k\to\infty}\mathbb{E}\left[|Y_k| : |Y_k| > m\right]$$

with the right member of this inequality going to zero as $m,M\to\infty$ with $M\ge m$ by the assumed uniform integrability of the $\{X_n\}$. Therefore there exists $X\in L^1(\Omega,\mathcal{B},P)$ such that $\lim_{m\to\infty}\mathbb{E}|X - X^m| = 0$.

We are now ready to verify Eq. (13.11) is valid. For $h\in\mathcal{B}_b$,

$$\begin{aligned}|\mathbb{E}[(X-Y_k)h]| &\le |\mathbb{E}[(X^m - Y_k^m)h]| + |\mathbb{E}[(X - X^m)h]| + |\mathbb{E}[(Y_k - Y_k^m)h]| \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot (\mathbb{E}[|X - X^m|] + \mathbb{E}[|Y_k| : |Y_k| > m]) \\ &\le |\mathbb{E}[(X^m - Y_k^m)h]| + \|h\|_\infty \cdot \left(\mathbb{E}[|X - X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right)\end{aligned}$$

Passing to the limit as $k\to\infty$ in the above inequality shows

$$\limsup_{k\to\infty} |\mathbb{E}[(X-Y_k)h]| \le \|h\|_\infty \cdot \left(\mathbb{E}[|X-X^m|] + \sup_l \mathbb{E}[|Y_l| : |Y_l| > m]\right).$$

Since $X^m \to X$ in L^1 and $\sup_l \mathbb{E}[|Y_l| : |Y_l| > m] \to 0$ by uniform integrability, it follows that, $\limsup_{k\to\infty} |\mathbb{E}[(X-Y_k)h]| = 0$. ■

Example 13.20. Let $(\Omega, \mathcal{B}, P) = \left((0,1), \mathcal{B}_{(0,1)}, m\right)$ where m is Lebesgue measure and let $X_n(\omega) = 2^n 1_{0<\omega<2^{-n}}$. Then $\mathbb{E}X_n = 1$ for all n and hence $\{X_n\}_{n=1}^\infty$ is bounded in $L^1(\Omega, \mathcal{B}, P)$ (but is not uniformly integrable). Suppose for sake of contradiction that there existed $X \in L^1(\Omega, \mathcal{B}, P)$ and subsequence, $Y_k := X_{n_k}$ such that $Y_k \stackrel{w}{\to} X$. Then for $h \in \mathcal{B}_b$ and any $\varepsilon > 0$ we would have

$$\mathbb{E}\left[Xh1_{(\varepsilon,1)}\right] = \lim_{k\to\infty} \mathbb{E}\left[Y_k h 1_{(\varepsilon,1)}\right] = 0.$$

Then by DCT it would follow that $\mathbb{E}[Xh] = 0$ for all $h \in \mathcal{B}_b$ and hence that $X \equiv 0$. On the other hand we would also have

$$0 = \mathbb{E}[X \cdot 1] = \lim_{k\to\infty} \mathbb{E}[Y_k \cdot 1] = 1$$

and we have reached the desired contradiction. Hence we must conclude that bounded subset of $L^1(\Omega, \mathcal{B}, P)$ need not be weakly compact and thus we can not drop the uniform integrability assumption made in Theorem 13.19.

When $1 < p < \infty$, the situation is simpler.

Theorem 13.21. *Let $p \in (1,\infty)$ and $q = p(p-1)^{-1} \in (1,\infty)$ be its conjugate exponent. If $\{X_n\}_{n=1}^\infty$ is a bounded sequence in $L^p(\Omega, \mathcal{B}, P)$, there exists $X \in L^p(\Omega, \mathcal{B}, P)$ and a subsequence $Y_k := X_{n_k}$ of $\{X_n\}_{n=1}^\infty$ such that*

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in L^q(\Omega, \mathcal{B}, P). \tag{13.12}$$

Proof. Let $C := \sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$ and recall that Lemma 12.40 guarantees that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$. Therefore by Theorem 13.19, there exists $X \in L^1(\Omega, \mathcal{B}, P)$ and a subsequence, $Y_k := X_{n_k}$, such that Eq. (13.11) holds. We will complete the proof by showing; a) $X \in L^p(\Omega, \mathcal{B}, P)$ and b) and Eq. (13.12) is valid.

a) For $h \in \mathcal{B}_b$ we have

$$|\mathbb{E}[Xh]| \le \liminf_{k\to\infty} \mathbb{E}[|Y_k h|] \le \liminf_{k\to\infty} \|Y_k\|_p \cdot \|h\|_q \le C\|h\|_q.$$

For $M < \infty$, taking $h = \overline{\operatorname{sgn}(X)}\, |X|^{p-1} 1_{|X|\le M}$ in the previous inequality shows

$$\begin{aligned} \mathbb{E}\left[|X|^p 1_{|X|\le M}\right] &\le C \left\| \overline{\operatorname{sgn}(X)}\, |X|^{p-1} 1_{|X|\le M} \right\|_q \\ &= C\left(\mathbb{E}\left[|X|^{(p-1)q} 1_{|X|\le M}\right]\right)^{1/q} \le C\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/q} \end{aligned}$$

from which it follows that

$$\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/p} \le \left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1-1/q} \le C.$$

Using the monotone convergence theorem, we may let $M \to \infty$ in this equation to find $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p} \le C < \infty$.

b) Now that we know $X \in L^p(\Omega, \mathcal{B}, P)$, in make sense to consider $\mathbb{E}[(X-Y_k)h]$ for all $h \in L^p(\Omega, \mathcal{B}, P)$. For $M < \infty$, let $h^M := h1_{|h|\le M}$, then

$$\begin{aligned} |\mathbb{E}[(X-Y_k)h]| &\le \left|\mathbb{E}\left[(X-Y_k)h^M\right]\right| + \left|\mathbb{E}\left[(X-Y_k)h1_{|h|>M}\right]\right| \\ &\le \left|\mathbb{E}\left[(X-Y_k)h^M\right]\right| + \|X-Y_k\|_p \left\|h1_{|h|>M}\right\|_q \\ &\le \left|\mathbb{E}\left[(X-Y_k)h^M\right]\right| + 2C\left\|h1_{|h|>M}\right\|_q. \end{aligned}$$

Since $h^M \in \mathcal{B}_b$, we may pass to the limit $k \to \infty$ in the previous inequality to find,

$$\limsup_{k\to\infty} |\mathbb{E}[(X-Y_k)h]| \le 2C\left\|h1_{|h|>M}\right\|_q.$$

This completes the proof, since $\left\|h1_{|h|>M}\right\|_q \to 0$ as $M \to \infty$ by DCT. ■

13.2 Exercises

Exercise 13.4. Suppose that $\{M_n\}_{n=1}^\infty$ is an increasing sequence of closed subspaces of a Hilbert space, H. Let M be the closure of $M_0 := \cup_{n=1}^\infty M_n$. Show $\lim_{n\to\infty} P_{M_n} x = P_M x$ for all $x \in H$. **Hint:** first prove this for $x \in M_0$ and then for $x \in M$. Also consider the case where $x \in M^\perp$.

Solution to Exercise (13.4). Let $P_n := P_{M_n}$ and $P = P_M$. If $y \in M_0$, then $P_n y = y = Py$ for all n sufficiently large. and therefore, $\lim_{n\to\infty} P_n y = Py$. Now suppose that $x \in M$ and $y \in M_0$. Then

$$\begin{aligned} \|Px - P_n x\| &\le \|Px - Py\| + \|Py - P_n y\| + \|P_n y - P_n x\| \\ &\le 2\|x-y\| + \|Py - P_n y\| \end{aligned}$$

and passing to the limit as $n \to \infty$ then shows

$$\limsup_{n\to\infty} \|Px - P_n x\| \le 2\|x-y\|.$$

The left hand side may be made as small as we like by choosing $y \in M_0$ arbitrarily close to $x \in M = \bar{M}_0$.

For the general case, if $x \in H$, then $x = Px + y$ where $y = x - Px \in M^\perp \subset M_n^\perp$ for all n. Therefore,

$$P_n x = P_n P x \to Px \text{ as } n \to \infty$$

by what we have just proved.

Exercise 13.5. Suppose that $(X, \|\cdot\|)$ is a normed space such that parallelogram law, Eq. (13.2), holds for all $x, y \in X$, then there exists a unique inner product on $\langle\cdot|\cdot\rangle$ such that $\|x\| := \sqrt{\langle x|x\rangle}$ for all $x \in X$. In this case we say that $\|\cdot\|$ is a Hilbertian norm.

Solution to Exercise (13.5). If $\|\cdot\|$ is going to come from an inner product $\langle\cdot|\cdot\rangle$, it follows from Eq. (13.1) that

$$2\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

and

$$-2\mathrm{Re}\langle x|y\rangle = \|x-y\|^2 - \|x\|^2 - \|y\|^2.$$

Subtracting these two equations gives the "polarization identity,"

$$4\mathrm{Re}\langle x|y\rangle = \|x+y\|^2 - \|x-y\|^2. \tag{13.13}$$

Replacing y by iy in this equation then implies that

$$4\mathrm{Im}\langle x|y\rangle = \|x+iy\|^2 - \|x-iy\|^2$$

from which we find

$$\langle x|y\rangle = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\varepsilon y\|^2 \tag{13.14}$$

where $G = \{\pm 1, \pm i\}$ – a cyclic subgroup of $S^1 \subset \mathbb{C}$. Hence, if $\langle\cdot|\cdot\rangle$ is going to exist we must define it by Eq. (13.14) and the uniqueness has been proved.

For existence, define $\langle x|y\rangle$ by Eq. (13.14) in which case,

$$\begin{aligned}\langle x|x\rangle &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\varepsilon x\|^2 = \frac{1}{4}\left[\|2x\|^2 + i\|x+ix\|^2 - i\|x-ix\|^2\right] \\ &= \|x\|^2 + \frac{i}{4}\left|1+i|^2\right|\|x\|^2 - \frac{i}{4}\left|1-i|^2\right|\|x\|^2 = \|x\|^2.\end{aligned}$$

So to finish the proof, it only remains to show that $\langle x|y\rangle$ defined by Eq. (13.14) is an inner product.

Since

$$\begin{aligned}4\langle y|x\rangle &= \sum_{\varepsilon\in G}\varepsilon\|y+\varepsilon x\|^2 = \sum_{\varepsilon\in G}\varepsilon\|\varepsilon\left(y+\varepsilon x\right)\|^2 \\ &= \sum_{\varepsilon\in G}\varepsilon\|\varepsilon y+\varepsilon^2 x\|^2 \\ &= \|y+x\|^2 - \|-y+x\|^2 + i\|iy-x\|^2 - i\|-iy-x\|^2 \\ &= \|x+y\|^2 - \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2 \\ &= 4\overline{\langle x|y\rangle}\end{aligned}$$

it suffices to show $x \to \langle x|y\rangle$ is linear for all $y \in H$. For this we will need to derive an identity from Eq. (13.2),. To do this we make use of Eq. (13.2), three times to find

$$\begin{aligned}\|x+y+z\|^2 &= -\|x+y-z\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|x-y-z\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= \|y+z-x\|^2 - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2 \\ &= -\|y+z+x\|^2 + 2\|y+z\|^2 + 2\|x\|^2 \\ &\quad - 2\|x-z\|^2 - 2\|y\|^2 + 2\|x+y\|^2 + 2\|z\|^2.\end{aligned}$$

Solving this equation for $\|x+y+z\|^2$ gives

$$\|x+y+z\|^2 = \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2. \tag{13.15}$$

Using Eq. (13.15), for $x, y, z \in H$,

$$\begin{aligned}4\,\mathrm{Re}\langle x+z|y\rangle &= \|x+z+y\|^2 - \|x+z-y\|^2 \\ &= \|y+z\|^2 + \|x+y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2 \\ &\quad - \left(\|z-y\|^2 + \|x-y\|^2 - \|x-z\|^2 + \|x\|^2 + \|z\|^2 - \|y\|^2\right) \\ &= \|z+y\|^2 - \|z-y\|^2 + \|x+y\|^2 - \|x-y\|^2 \\ &= 4\,\mathrm{Re}\langle x|y\rangle + 4\,\mathrm{Re}\langle z|y\rangle.\end{aligned} \tag{13.16}$$

Now suppose that $\delta \in G$, then since $|\delta| = 1$,

$$\begin{aligned}4\langle\delta x|y\rangle &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|\delta x+\varepsilon y\|^2 = \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\|x+\delta^{-1}\varepsilon y\|^2 \\ &= \frac{1}{4}\sum_{\varepsilon\in G}\varepsilon\delta\|x+\delta\varepsilon y\|^2 = 4\delta\langle x|y\rangle\end{aligned} \tag{13.17}$$

where in the third inequality, the substitution $\varepsilon \to \varepsilon\delta$ was made in the sum. So Eq. (13.17) says $\langle \pm ix|y\rangle = \pm i\langle x|y\rangle$ and $\langle -x|y\rangle = -\langle x|y\rangle$. Therefore

$$\mathrm{Im}\langle x|y\rangle = \mathrm{Re}\left(-i\langle x|y\rangle\right) = \mathrm{Re}\langle -ix|y\rangle$$

which combined with Eq. (13.16) shows

$$\begin{aligned}\mathrm{Im}\langle x+z|y\rangle &= \mathrm{Re}\langle -ix-iz|y\rangle = \mathrm{Re}\langle -ix|y\rangle + \mathrm{Re}\langle -iz|y\rangle \\ &= \mathrm{Im}\langle x|y\rangle + \mathrm{Im}\langle z|y\rangle\end{aligned}$$

and therefore (again in combination with Eq. (13.16)),

$$\langle x+z|y\rangle = \langle x|y\rangle + \langle z|y\rangle \text{ for all } x, y \in H.$$

Because of this equation and Eq. (13.17) to finish the proof that $x \to \langle x|y\rangle$ is linear, it suffices to show $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$. Now if $\lambda = m \in \mathbb{N}$, then

$$\langle mx|y\rangle = \langle x + (m-1)x|y\rangle = \langle x|y\rangle + \langle (m-1)x|y\rangle$$

so that by induction $\langle mx|y\rangle = m\langle x|y\rangle$. Replacing x by x/m then shows that $\langle x|y\rangle = m\langle m^{-1}x|y\rangle$ so that $\langle m^{-1}x|y\rangle = m^{-1}\langle x|y\rangle$ and so if $m, n \in \mathbb{N}$, we find

$$\langle \frac{n}{m}x|y\rangle = n\langle \frac{1}{m}x|y\rangle = \frac{n}{m}\langle x|y\rangle$$

so that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$ and $\lambda \in \mathbb{Q}$. By continuity, it now follows that $\langle \lambda x|y\rangle = \lambda\langle x|y\rangle$ for all $\lambda > 0$.

An alternate ending: In the case where X is real, the latter parts of the proof are easier to digest as we can use Eq. (13.13) for the formula for the inner product. For example, we have

$$\begin{aligned}4\langle x|2z\rangle &= \|x+2z\|^2 - \|x-2z\|^2 \\ &= \|x+z+z\|^2 + \|x+z-z\|^2 - \|x-z+z\|^2 - \|x-z-z\|^2 \\ &= \frac{1}{2}\left[\|x+z\|^2 + \|z\|^2\right] - \frac{1}{2}\left[\|x-z\|^2 + \|z\|^2\right] \\ &= \frac{1}{2}\left[\|x+z\|^2 - \|x-z\|^2\right] = 2\langle x|z\rangle\end{aligned}$$

from which it follows that $\langle x|2z\rangle = 2\langle x|z\rangle$. Similarly,

$$\begin{aligned}4[\langle x|z\rangle + \langle y|z\rangle] &= \|x+z\|^2 - \|x-z\|^2 + \|y+z\|^2 - \|y-z\|^2 \\ &= \|x+z\|^2 + \|y+z\|^2 - \|x-z\|^2 - \|y-z\|^2 \\ &= \frac{1}{2}\left(\|x+y+2z\|^2 + \|x-y\|^2\right) - \frac{1}{2}\left(\|x+y-2z\|^2 + \|x-y\|^2\right) \\ &= 2\langle x+y|2z\rangle = 4\langle x+y|z\rangle\end{aligned}$$

from which it follows that and $\langle x+y|z\rangle = \langle x|z\rangle + \langle y|z\rangle$. From this identity one shows as above that $\langle\cdot|\cdot\rangle$ is a real inner product on X.

Now suppose that X is complex and now let

$$Q(x,y) = \frac{1}{4}\left[\|x+z\|^2 - \|x-z\|^2\right].$$

We should expect that $Q(\cdot,\cdot) = \mathrm{Re}\langle\cdot|\cdot\rangle$ and therefore we should define

$$\langle x|y\rangle := Q(x,y) - iQ(ix,y).$$

Since

$$\begin{aligned}4Q(ix,y) &= \|ix+y\|^2 - \|ix-y\|^2 = \|-i(ix+y)\|^2 - \|-i(ix-y)\|^2 \\ &= \|x-iy\|^2 - \|x+iy\|^2 = -4Q(x,iy),\end{aligned}$$

it follows that $Q(ix,x) = 0$ so that $\langle x|x\rangle = \|x\|^2$ and that

$$\langle y|x\rangle = Q(y,x) - iQ(iy,x) = Q(y,x) + iQ(y,ix) = \overline{\langle x|y\rangle}.$$

Since $x \to \langle x|y\rangle$ is real linear, we now need only show that $\langle ix|y\rangle = i\langle x|y\rangle$. However,

$$\begin{aligned}\langle ix|y\rangle &= Q(ix,y) - iQ(i(ix),y) \\ &= Q(ix,y) + iQ(x,y) = i\langle x|y\rangle\end{aligned}$$

as desired.

Conditional Expectation

In this section let $(\Omega,\mathcal{B},P)$ be a probability space and $\mathcal{G}\subset\mathcal{B}$ be a sub – sigma algebra of $\mathcal{B}$. We will write $f\in\mathcal{G}_b$ iff $f:\Omega\to\mathbb{C}$ is bounded and f is $(\mathcal{G},\mathcal{B}_{\mathbb{C}})$ – measurable. If $A\in\mathcal{B}$ and $P(A)>0$, we will let

$$\mathbb{E}[X|A]:=\frac{\mathbb{E}[X:A]}{P(A)}\text{ and }P(B|A):=\mathbb{E}[1_B|A]:=\frac{P(A\cap B)}{P(A)}$$

for all integrable random variables, X, and $B\in\mathcal{B}$. We will often use the factorization Lemma 6.40 in this section. Because of this let us repeat it here.

Lemma 14.1. *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 $H=h\circ Y$.*

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 $H=\sum a_i1_{A_i}$ 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 $h:=\sum a_i1_{B_i}$ – 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 $\bar{\mathbb{R}}$. ■

Definition 14.2 (Conditional Expectation). *Let $\mathbb{E}_{\mathcal{G}}:L^2(\Omega,\mathcal{B},P)\to L^2(\Omega,\mathcal{G},P)$ denote orthogonal projection of $L^2(\Omega,\mathcal{B},P)$ onto the closed subspace $L^2(\Omega,\mathcal{G},P)$. For $f\in L^2(\Omega,\mathcal{B},P)$, we say that $\mathbb{E}_{\mathcal{G}}f\in L^2(\Omega,\mathcal{G},P)$ is the* ***conditional expectation*** *of f.*

Remark 14.3 (Basic Properties of $\mathbb{E}_{\mathcal{G}}$). Let $f\in L^2(\Omega,\mathcal{B},P)$. By the orthogonal projection Theorem 13.12 we know that $F\in L^2(\Omega,\mathcal{G},P)$ is $\mathbb{E}_{\mathcal{G}}f$ a.s. iff either of the following two conditions hold;

1. $\|f-F\|_2\le\|f-g\|_2$ for all $g\in L^2(\Omega,\mathcal{G},P)$ or
2. $\mathbb{E}[fh]=\mathbb{E}[Fh]$ for all $h\in L^2(\Omega,\mathcal{G},P)$.

Moreover if $\mathcal{G}_0\subset\mathcal{G}_1\subset\mathcal{B}$ then $L^2(\Omega,\mathcal{G}_0,P)\subset L^2(\Omega,\mathcal{G}_1,P)\subset L^2(\Omega,\mathcal{B},P)$ and therefore,

$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}f=\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0}f=\mathbb{E}_{\mathcal{G}_0}f\text{ a.s. for all }f\in L^2(\Omega,\mathcal{B},P).\tag{14.1}$$

It is also useful to observe that condition 2. above may expressed as

$$\mathbb{E}[f:A]=\mathbb{E}[F:A]\text{ for all }A\in\mathcal{G}\tag{14.2}$$

or

$$\mathbb{E}[fh]=\mathbb{E}[Fh]\text{ for all }h\in\mathcal{G}_b.\tag{14.3}$$

Indeed, if Eq. (14.2) holds, then by linearity we have $\mathbb{E}[fh]=\mathbb{E}[Fh]$ for all $\mathcal{G}$ – measurable simple functions, h and hence by the approximation Theorem 6.39 and the DCT for all $h\in\mathcal{G}_b$. Therefore Eq. (14.2) implies Eq. (14.3). If Eq. (14.3) holds and $h\in L^2(\Omega,\mathcal{G},P)$, we may use DCT to show

$$\mathbb{E}[fh]\overset{\text{DCT}}{=}\lim_{n\to\infty}\mathbb{E}\left[fh1_{|h|\le n}\right]\overset{(14.3)}{=}\lim_{n\to\infty}\mathbb{E}\left[Fh1_{|h|\le n}\right]\overset{\text{DCT}}{=}\mathbb{E}[Fh],$$

which is condition 2. in Remark 14.3. Taking $h=1_A$ with $A\in\mathcal{G}$ in condition 2. or Remark 14.3, we learn that Eq. (14.2) is satisfied as well.

Theorem 14.4. *Let $(\Omega,\mathcal{B},P)$ and $\mathcal{G}\subset\mathcal{B}$ be as above and let $f,g\in L^1(\Omega,\mathcal{B},P)$. The operator $\mathbb{E}_{\mathcal{G}}:L^2(\Omega,\mathcal{B},P)\to L^2(\Omega,\mathcal{G},P)$ extends uniquely to a linear contraction from $L^1(\Omega,\mathcal{B},P)$ to $L^1(\Omega,\mathcal{G},P)$. This extension enjoys the following properties;*

1. *If $f\ge0$, P – a.e. then $\mathbb{E}_{\mathcal{G}}f\ge0$, P – a.e.*
2. ***Monotonicity.*** *If $f\ge g$, P – a.e. there $\mathbb{E}_{\mathcal{G}}f\ge\mathbb{E}_{\mathcal{G}}g$, P – a.e.*
3. *L^∞* ***– contraction property.*** *$|\mathbb{E}_{\mathcal{G}}f|\le\mathbb{E}_{\mathcal{G}}|f|$, P – a.e.*
4. ***Averaging Property.*** *If $f\in L^1(\Omega,\mathcal{B},P)$ then $F=\mathbb{E}_{\mathcal{G}}f$ iff $F\in L^1(\Omega,\mathcal{G},P)$ and*
$$\mathbb{E}(Fh)=\mathbb{E}(fh)\text{ for all }h\in\mathcal{G}_b.\tag{14.4}$$
5. ***Pull out property or product rule.*** *If $g\in\mathcal{G}_b$ and $f\in L^1(\Omega,\mathcal{B},P)$, then $\mathbb{E}_{\mathcal{G}}(gf)=g\cdot\mathbb{E}_{\mathcal{G}}f$, P – a.e.*

6. ***Tower or smoothing property.*** *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*

$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^1(\Omega, \mathcal{B}, P). \tag{14.5}$$

Proof. By the definition of orthogonal projection, $f \in L^2(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$,

$$\mathbb{E}(fh) = \mathbb{E}(f \cdot \mathbb{E}_{\mathcal{G}} h) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h). \tag{14.6}$$

Taking

$$h = \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} := \frac{\overline{\mathbb{E}_{\mathcal{G}} f}}{\mathbb{E}_{\mathcal{G}} f} 1_{|\mathbb{E}_{\mathcal{G}} f| > 0} \tag{14.7}$$

in Eq. (14.6) shows

$$\mathbb{E}(|\mathbb{E}_{\mathcal{G}} f|) = \mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \le \mathbb{E}(|fh|) \le \mathbb{E}(|f|). \tag{14.8}$$

It follows from this equation and the BLT (Theorem 13.17) that $\mathbb{E}_{\mathcal{G}}$ extends uniquely to a contraction form $L^1(\Omega, \mathcal{B}, P)$ to $L^1(\Omega, \mathcal{G}, P)$. Moreover, by a simple limiting argument, Eq. (14.6) remains valid for all $f \in L^1(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$. Indeed, (without reference to Theorem 13.17) if $f_n := f 1_{|f| \le n} \in L^2(\Omega, \mathcal{B}, P)$, then $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence

$$\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f_n - \mathbb{E}_{\mathcal{G}} f_m|] = \mathbb{E}[|\mathbb{E}_{\mathcal{G}}(f_n - f_m)|] \le \mathbb{E}[|f_n - f_m|] \to 0 \text{ as } m, n \to \infty.$$

By the completeness of $L^1(\Omega, \mathcal{G}, P)$, $F := L^1(\Omega, \mathcal{G}, P)\text{-}\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n$ exists. Moreover the function F satisfies,

$$\mathbb{E}(F \cdot h) = \mathbb{E}(\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n \cdot h) = \lim_{n\to\infty} \mathbb{E}(f_n \cdot h) = \mathbb{E}(f \cdot h) \tag{14.9}$$

for all $h \in \mathcal{G}_b$ and by Proposition 7.22 there is at most one, $F \in L^1(\Omega, \mathcal{G}, P)$, which satisfies Eq. (14.9). We will again denote F by $\mathbb{E}_{\mathcal{G}} f$. This proves the existence and uniqueness of F satisfying the defining relation in Eq. (14.4) of item 4. The same argument used in Eq. (14.8) again shows $\mathbb{E}|F| \le \mathbb{E}|f|$ and therefore that $\mathbb{E}_{\mathcal{G}} : L^1(\Omega, \mathcal{B}, P) \to L^1(\Omega, \mathcal{G}, P)$ is a contraction.

Items 1 and 2. If $f \in L^1(\Omega, \mathcal{B}, P)$ with $f \ge 0$, then

$$\mathbb{E}(\mathbb{E}_{\mathcal{G}} f \cdot h) = \mathbb{E}(fh) \ge 0 \ \forall\, h \in \mathcal{G}_b \text{ with } h \ge 0. \tag{14.10}$$

An application of Lemma 7.24 then shows that $\mathbb{E}_{\mathcal{G}} f \ge 0$ a.s.[1] The proof of item 2. follows by applying item 1. with f replaced by $f - g \ge 0$.

Item 3. If f is real, $\pm f \le |f|$ and so by Item 2., $\pm\mathbb{E}_{\mathcal{G}} f \le \mathbb{E}_{\mathcal{G}} |f|$, i.e. $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, P – a.e. For complex f, let $h \ge 0$ be a bounded and $\mathcal{G}$ – measurable function. Then

[1] This can also easily be proved directly here by taking $h = 1_{\mathbb{E}_{\mathcal{G}} f < 0}$ in Eq. (14.10).

$$\begin{aligned}\mathbb{E}[|\mathbb{E}_{\mathcal{G}} f| h] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{G}} f \cdot \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right] = \mathbb{E}\left[f \cdot \overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}} f)} h\right] \\ &\le \mathbb{E}[|f| h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} |f| \cdot h].\end{aligned}$$

Since $h \ge 0$ is an arbitrary $\mathcal{G}$ – measurable function, it follows, by Lemma 7.24, that $|\mathbb{E}_{\mathcal{G}} f| \le \mathbb{E}_{\mathcal{G}} |f|$, P – a.s. Recall the item 4. has already been proved.

Item 5. If $h, g \in \mathcal{G}_b$ and $f \in L^1(\Omega, \mathcal{B}, P)$, then

$$\mathbb{E}[(g\mathbb{E}_{\mathcal{G}} f) h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}} f \cdot hg] = \mathbb{E}[f \cdot hg] = \mathbb{E}[gf \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}}(gf) \cdot h].$$

Thus $\mathbb{E}_{\mathcal{G}}(gf) = g \cdot \mathbb{E}_{\mathcal{G}} f$, P – a.e.

Item 6., by the item 5. of the projection Theorem 13.12, Eq. (14.5) holds on $L^2(\Omega, \mathcal{B}, P)$. By continuity of conditional expectation on $L^1(\Omega, \mathcal{B}, P)$ and the density of L^1 probability spaces in L^2 – probability spaces shows that Eq. (14.5) continues to hold on $L^1(\Omega, \mathcal{B}, P)$.

Second Proof. For $h \in (\mathcal{G}_0)_b$, we have

$$\mathbb{E}[\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_1} f \cdot h] = \mathbb{E}[f \cdot h] = \mathbb{E}[\mathbb{E}_{\mathcal{G}_0} f \cdot h]$$

which shows $\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_0} f$ a.s. By the product rule in item 5., it also follows that

$$\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f] = \mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f \cdot 1] = \mathbb{E}_{\mathcal{G}_0} f \cdot \mathbb{E}_{\mathcal{G}_1}[1] = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s.}$$

Notice that $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ need only be $\mathcal{G}_1$ – measurable. What the statement says there are representatives of $\mathbb{E}_{\mathcal{G}_1}[\mathbb{E}_{\mathcal{G}_0} f]$ which is $\mathcal{G}_0$ – measurable and any such representative is also a representative of $\mathbb{E}_{\mathcal{G}_0} f$. ■

Remark 14.5. There is another standard construction of $\mathbb{E}_{\mathcal{G}} f$ based on the characterization in Eq. (14.4) and the Radon Nikodym Theorem 15.8. It goes as follows, for $0 \le f \in L^1(P)$, let $Q := fP$ and observe that $Q|_{\mathcal{G}} \ll P|_{\mathcal{G}}$ and hence there exists $0 \le g \in L^1(\Omega, \mathcal{G}, P)$ such that $dQ|_{\mathcal{G}} = g dP|_{\mathcal{G}}$. This then implies that

$$\int_A f dP = Q(A) = \int_A g dP \text{ for all } A \in \mathcal{G},$$

i.e. $g = \mathbb{E}_{\mathcal{G}} f$. For general real valued, $f \in L^1(P)$, define $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} f_+ - \mathbb{E}_{\mathcal{G}} f_-$ and then for complex $f \in L^1(P)$ let $\mathbb{E}_{\mathcal{G}} f = \mathbb{E}_{\mathcal{G}} \operatorname{Re} f + i\mathbb{E}_{\mathcal{G}} \operatorname{Im} f$.

Notation 14.6 *In the future, we will often write $\mathbb{E}_{\mathcal{G}} f$ as $\mathbb{E}[f|\mathcal{G}]$. Moreover, if $(\mathbb{X}, \mathcal{M})$ is a measurable space and $X : \Omega \to \mathbb{X}$ is a measurable map. We will often simply denote $\mathbb{E}[f|\sigma(X)]$ simply by $\mathbb{E}[f|X]$. We will further let $P(A|\mathcal{G}) := \mathbb{E}[1_A|\mathcal{G}]$ be the* ***conditional probability of*** *A* ***given*** *$\mathcal{G}$, and $P(A|X) := P(A|\sigma(X))$ be* ***conditional probability of*** *A* ***given*** *X.*

Exercise 14.1. Suppose $f \in L^1(\Omega, \mathcal{B}, P)$ and $f > 0$ a.s. Show $\mathbb{E}[f|\mathcal{G}] > 0$ a.s. Use this result to conclude if $f \in (a, b)$ a.s. for some a, b such that $-\infty \le a < b \le \infty$, then $\mathbb{E}[f|\mathcal{G}] \in (a, b)$ a.s. More precisely you are to show that any version, g, of $\mathbb{E}[f|\mathcal{G}]$ satisfies, $g \in (a, b)$ a.s.

14.1 Examples

Example 14.7. Suppose $\mathcal{G}$ is the trivial $\sigma-$algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$. In this case $\mathbb{E}_\mathcal{G} f = \mathbb{E} f$ a.s.

Example 14.8. On the opposite extreme, if $\mathcal{G} = \mathcal{B}$, then $\mathbb{E}_\mathcal{G} f = f$ a.s.

Lemma 14.9. *Suppose $(\mathbb{X}, \mathcal{M})$ is a measurable space, $X : \Omega \to \mathbb{X}$ is a measurable function, and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{B}$. If X is independent of $\mathcal{G}$ and $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_\mathcal{G}[f(X)] = \mathbb{E}[f(X)]$ a.s.. Conversely if $\mathbb{E}_\mathcal{G}[f(X)] = \mathbb{E}[f(X)]$ a.s. for all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, then X is independent of $\mathcal{G}$.*

Proof. Suppose that X is independent of $\mathcal{G}$, $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $f(X) \in L(\Omega, \mathcal{B}, P)$, $\mu := \mathbb{E}[f(X)]$, and $A \in \mathcal{G}$. Then, by independence,

$$\mathbb{E}[f(X) : A] = \mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A] = \mathbb{E}[\mu 1_A] = \mathbb{E}[\mu : A].$$

Therefore $\mathbb{E}_\mathcal{G}[f(X)] = \mu = \mathbb{E}[f(X)]$ a.s.

Conversely if $\mathbb{E}_\mathcal{G}[f(X)] = \mathbb{E}[f(X)] = \mu$ and $A \in \mathcal{G}$, then

$$\mathbb{E}[f(X) 1_A] = \mathbb{E}[f(X) : A] = \mathbb{E}[\mu : A] = \mu \mathbb{E}[1_A] = \mathbb{E}[f(X)] \mathbb{E}[1_A].$$

Since this last equation is assumed to hold true for all $A \in \mathcal{G}$ and all bounded measurable functions, $f : \mathbb{X} \to \mathbb{R}$, X is independent of $\mathcal{G}$. ∎

The following remark is often useful in computing conditional expectations. The following Exercise should help you gain some more intuition about conditional expectations.

*Remark 14.10 (**Note well.**).* According to Lemma 14.1, $\mathbb{E}(f|X) = \tilde{f}(X)$ a.s. for some measurable function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$. So computing $\mathbb{E}(f|X) = \tilde{f}(X)$ is equivalent to finding a function, $\tilde{f} : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f \cdot h(X)] = \mathbb{E}\left[\tilde{f}(X) h(X)\right] \tag{14.11}$$

for all bounded and measurable functions, $h : \mathbb{X} \to \mathbb{R}$.

Exercise 14.2. Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and $\mathcal{P} := \{A_i\}_{i=1}^\infty \subset \mathcal{B}$ is a partition of Ω. (Recall this means $\Omega = \sum_{i=1}^\infty A_i$.) Let $\mathcal{G}$ be the $\sigma-$algebra generated by $\mathcal{P}$. Show:

1. $B \in \mathcal{G}$ iff $B = \cup_{i \in \Lambda} A_i$ for some $\Lambda \subset \mathbb{N}$.
2. $g : \Omega \to \mathbb{R}$ is $\mathcal{G}-$measurable iff $g = \sum_{i=1}^\infty \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$.
3. For $f \in L^1(\Omega, \mathcal{B}, P)$, let $\mathbb{E}[f|A_i] := \mathbb{E}[1_{A_i} f] / P(A_i)$ if $P(A_i) \neq 0$ and $\mathbb{E}[f|A_i] = 0$ otherwise. Show

$$\mathbb{E}_\mathcal{G} f = \sum_{i=1}^\infty \mathbb{E}[f|A_i] 1_{A_i} \text{ a.s.} \tag{14.12}$$

Solution to Exercise (14.2). We will only prove part 3. here. To do this, suppose that $\mathbb{E}_\mathcal{G} f = \sum_{i=1}^\infty \lambda_i 1_{A_i}$ for some $\lambda_i \in \mathbb{R}$. Then

$$\mathbb{E}[f : A_j] = \mathbb{E}[\mathbb{E}_\mathcal{G} f : A_j] = \mathbb{E}\left[\sum_{i=1}^\infty \lambda_i 1_{A_i} : A_j\right] = \lambda_j P(A_j)$$

which holds automatically if $P(A_j) = 0$ no matter how λ_j is chosen. Therefore, we must take

$$\lambda_j = \frac{\mathbb{E}[f : A_j]}{P(A_j)} = \mathbb{E}[f|A_j]$$

which verifies Eq. (14.12).

Proposition 14.11. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, $(\mathbb{X}, \mathcal{M}, \mu)$ and $(\mathbb{Y}, \mathcal{N}, \nu)$ are two σ – finite measure spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions, and there exists $0 \le \rho \in L^1(\Omega, \mathcal{B}, \mu \otimes \nu)$ such that $P((X, Y) \in U) = \int_U \rho(x, y) \, d\mu(x) \, d\nu(y)$ for all $U \in \mathcal{M} \otimes \mathcal{N}$. Let*

$$\bar{\rho}(x) := \int_\mathbb{Y} \rho(x, y) \, d\nu(y) \tag{14.13}$$

and $x \in \mathbb{X}$ and $B \in \mathcal{N}$, let

$$Q(x, B) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_B \rho(x, y) \, d\nu(y) & \text{if } \bar{\rho}(x) \in (0, \infty) \\ \delta_{y_0}(B) & \text{if } \bar{\rho}(x) \in \{0, \infty\} \end{cases} \tag{14.14}$$

where y_0 is some arbitrary but fixed point in Y. Then for any bounded (or non-negative) measurable function, $f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, we have

$$\mathbb{E}[f(X, Y) | X] = Q(X, f(X, \cdot)) =: \int_\mathbb{Y} f(X, y) \, Q(X, dy) = g(X) \;\; a.s. \tag{14.15}$$

where,

$$g(x) := \int_\mathbb{Y} f(x, y) \, Q(x, dy) = Q(x, f(x, \cdot)).$$

As usual we use the notation,

$$Q(x, v) := \int_\mathbb{Y} v(y) \, Q(x, dy) = \begin{cases} \frac{1}{\bar{\rho}(x)} \int_\mathbb{Y} v(y) \rho(x, y) \, d\nu(y) & \text{if } \bar{\rho}(x) \in (0, \infty) \\ \delta_{y_0}(v) = v(y_0) & \text{if } \bar{\rho}(x) \in \{0, \infty\}. \end{cases}$$

for all bounded measurable functions, $v : \mathbb{Y} \to \mathbb{R}$,

Proof. Our goal is to compute $\mathbb{E}[f(X,Y)|X]$. According to Remark 14.10, we are searching for a bounded measurable function, $g:\mathbb{X}\to\mathbb{R}$, such that

$$\mathbb{E}[f(X,Y)h(X)]=\mathbb{E}[g(X)h(X)] \text{ for all } h\in\mathcal{M}_b. \tag{14.16}$$

(Throughout this argument we are going to repeatedly use the Tonelli - Fubini theorems.) We now explicitly write out both sides of Eq. (14.16);

$$\begin{aligned}\mathbb{E}[f(X,Y)h(X)]&=\int_{\mathbb{X}\times\mathbb{Y}}h(x)f(x,y)\rho(x,y)\,d\mu(x)\,d\nu(y)\\&=\int_{\mathbb{X}}h(x)\left[\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)\right]d\mu(x)\end{aligned} \tag{14.17}$$

$$\begin{aligned}\mathbb{E}[g(X)h(X)]&=\int_{\mathbb{X}\times\mathbb{Y}}h(x)g(x)\rho(x,y)\,d\mu(x)\,d\nu(y)\\&=\int_{\mathbb{X}}h(x)g(x)\bar\rho(x)\,d\mu(x).\end{aligned} \tag{14.18}$$

Since the right sides of Eqs. (14.17) and (14.18) must be equal for all $h\in\mathcal{M}_b$, we must demand,

$$\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)=g(x)\bar\rho(x) \text{ for } \mu-\text{a.e. } x. \tag{14.19}$$

There are two possible problems in solving this equation for $g(x)$ at a particular point x; the first is when $\bar\rho(x)=0$ and the second is when $\bar\rho(x)=\infty$. Since

$$\int_{\mathbb{X}}\bar\rho(x)\,d\mu(x)=\int_{\mathbb{X}}\left[\int_{\mathbb{Y}}\rho(x,y)\,d\nu(y)\right]d\mu(x)=1,$$

we know that $\bar\rho(x)<\infty$ for $\mu-$ a.e. x and therefore

$$P(X\in\{\bar\rho=0\})=P(\bar\rho(X)=0)=\int_{\mathbb{X}}1_{\bar\rho=0}\bar\rho\,d\mu=0.$$

Hence the points where $\bar\rho(x)=\infty$ will not cause any problems.

For the first problem, namely points x where $\bar\rho(x)=0$, we know that $\rho(x,y)=0$ for $\nu-$ a.e. y and therefore

$$\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)=0. \tag{14.20}$$

Hence at such points, x where $\bar\rho(x)=0$, Eq. (14.19) will be valid no matter how we choose $g(x)$. Therefore, if we let $y_0\in\mathbb{Y}$ be an arbitrary but fixed point and then define

$$g(x):=\begin{cases}\frac{1}{\bar\rho(x)}\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y) & \text{if } \bar\rho(x)\in(0,\infty)\\ f(x,y_0) & \text{if } \bar\rho(x)\in\{0,\infty\}\end{cases},$$

then we have shown $\mathbb{E}[f(X,Y)|X]=g(X)=Q(X,f)$ a.s. as desired. (Observe here that when $\bar\rho(x)<\infty$, $\rho(x,\cdot)\in L^1(\nu)$ and hence the integral in the definition of g is well defined.)

Just for added security, let us check directly that $g(X)=\mathbb{E}[f(X,Y)|X]$ a.s.. According to Eq. (14.18) we have

$$\begin{aligned}\mathbb{E}[g(X)h(X)]&=\int_{\mathbb{X}}h(x)g(x)\bar\rho(x)\,d\mu(x)\\&=\int_{\mathbb{X}\cap\{0<\bar\rho<\infty\}}h(x)g(x)\bar\rho(x)\,d\mu(x)\\&=\int_{\mathbb{X}\cap\{0<\bar\rho<\infty\}}h(x)\bar\rho(x)\left(\frac{1}{\bar\rho(x)}\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x)\\&=\int_{\mathbb{X}\cap\{0<\bar\rho<\infty\}}h(x)\left(\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x)\\&=\int_{\mathbb{X}}h(x)\left(\int_{\mathbb{Y}}f(x,y)\rho(x,y)\,d\nu(y)\right)d\mu(x)\\&=\mathbb{E}[f(X,Y)h(X)]\qquad\text{(by Eq. (14.17))},\end{aligned}$$

wherein we have repeatedly used $\mu(\bar\rho=\infty)=0$ and Eq. (14.20) holds when $\bar\rho(x)=0$. This completes the verification that $g(X)=\mathbb{E}[f(X,Y)|X]$ a.s.. ■

This proposition shows that conditional expectation is a generalization of the notion of performing integration over a partial subset of the variables in the integrand. Whereas to compute the expectation, one should integrate over all of the variables. It also gives an example of regular conditional probabilities.

Definition 14.12. *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces. A function, $Q:\mathbb{X}\times\mathcal{N}\to[0,1]$ is a* ***probability kernel on*** *$\mathbb{X}\times\mathbb{Y}$ iff*

1. *$Q(x,\cdot):\mathcal{N}\to[0,1]$ is a probability measure on $(\mathbb{Y},\mathcal{N})$ for each $x\in\mathbb{X}$ and*
2. *$Q(\cdot,B):\mathbb{X}\to[0,1]$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable for all $B\in\mathcal{N}$.*

If Q is a probability kernel on $\mathbb{X}\times\mathbb{Y}$ and $f:\mathbb{Y}\to\mathbb{R}$ is a bounded measurable function or a positive measurable function, then $x\to Q(x,f):=\int_{\mathbb{Y}}f(y)Q(x,dy)$ is $\mathcal{M}/\mathcal{B}_{\mathbb{R}}$ – measurable. This is clear for simple functions and then for general functions via simple limiting arguments.

Definition 14.13. *Let $(\mathbb{X},\mathcal{M})$ and $(\mathbb{Y},\mathcal{N})$ be measurable spaces and $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ be measurable functions. A probability kernel, Q, on $\mathbb{X}\times\mathbb{Y}$ is said to be a* ***regular conditional distribution of*** *Y* ***given*** *X iff $Q(X,B)$ is a version of $P(Y\in B|X)$ for each $B\in\mathcal{N}$. Equivalently, we should have*

$Q(X, f) = \mathbb{E}[f(Y)|X]$ *a.s. for all* $f \in \mathcal{N}_b$. *When* $\mathbb{X} = \Omega$ *and* $\mathcal{M} = \mathcal{G}$ *is a sub-*σ *– algebra of* $\mathcal{B}$, *we say that* Q *is the* ***regular conditional distribution of*** Y ***given*** $\mathcal{G}$.

The probability kernel, Q, defined in Eq. (14.14) is an example of a regular conditional distribution of Y given X. In general if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{B}$. Letting $P_{\mathcal{G}}(A) = P(A|\mathcal{G}) := \mathbb{E}[1_A|\mathcal{G}] \in L^2(\Omega, \mathcal{B}, P)$ for all $A \in \mathcal{B}$, then $P_{\mathcal{G}} : \mathcal{B} \to L^2(\Omega, \mathcal{G}, P)$ is a map such that whenever $A, A_n \in \mathcal{B}$ with $A = \sum_{n=1}^{\infty} A_n$, we have (by cDCT) that

$$P_{\mathcal{G}}(A) = \sum_{n=1}^{\infty} P_{\mathcal{G}}(A_n) \text{ (equality in } L^2(\Omega, \mathcal{G}, P). \tag{14.21}$$

Now suppose that we have chosen a representative, $\bar{P}_{\mathcal{G}}(A) : \Omega \to [0,1]$, of $P_{\mathcal{G}}(A)$ for each $A \in \mathcal{B}$. From Eq. (14.21) it follows that

$$\bar{P}_{\mathcal{G}}(A)(\omega) = \sum_{n=1}^{\infty} \bar{P}_{\mathcal{G}}(A_n)(\omega) \text{ for } P\text{-a.e. } \omega. \tag{14.22}$$

However, **note well**, the exceptional set of ω's depends on the sets $A, A_n \in \mathcal{B}$. The goal of regular conditioning is to carefully choose the representative, $\bar{P}_{\mathcal{G}}(A) : \Omega \to [0,1]$, such that Eq. (14.22) holds for all $\omega \in \Omega$ and all $A, A_n \in \mathcal{B}$ with $A = \sum_{n=1}^{\infty} A_n$.

Remark 14.14. Unfortunately, regular conditional distributions do not always exists. However, if we require $\mathbb{Y}$ to be a "standard Borel space," (i.e. $\mathbb{Y}$ is isomorphic to a Borel subset of $\mathbb{R}$), then a conditional distribution of Y given X will always exists. See Theorem 14.24. Moreover, it is known that all "reasonable" measure spaces are standard Borel spaces, see Section 9.10 below for more details. So in most instances of interest a regular conditional distribution of Y given X **will** exist.

Exercise 14.3. Suppose that $(\mathbb{X}, \mathcal{M})$ and $(\mathbb{Y}, \mathcal{N})$ are measurable spaces, $X : \Omega \to \mathbb{X}$ and $Y : \Omega \to \mathbb{Y}$ are measurable functions, and there exists a regular conditional distribution, Q, of Y given X. Show:

1. For all bounded measurable functions, $f : (\mathbb{X} \times \mathbb{Y}, \mathcal{M} \otimes \mathcal{N}) \to \mathbb{R}$, the function $\mathbb{X} \ni x \to Q(x, f(x, \cdot))$ is measurable and
$$Q(X, f(X, \cdot)) = \mathbb{E}[f(X, Y)|X] \text{ a.s.} \tag{14.23}$$
Hint: let $\mathbb{H}$ denote the set of bounded measurable functions, f, on $\mathbb{X} \times \mathbb{Y}$ such that the two assertions are valid.
2. If $A \in \mathcal{M} \otimes \mathcal{N}$ and $\mu := P \circ X^{-1}$ be the law of X, then
$$P((X, Y) \in A) = \int_{\mathbb{X}} Q(x, 1_A(x, \cdot))\, d\mu(x) = \int_{\mathbb{X}} d\mu(x) \int_{\mathbb{Y}} 1_A(x, y)\, Q(x, dy). \tag{14.24}$$

Exercise 14.4. Keeping the same notation as in Exercise 14.3 and further assume that X and Y are independent. Find a regular conditional distribution of Y given X and prove

$$\mathbb{E}[f(X, Y)|X] = h_f(X) \text{ a.s. } \forall \text{ bounded measurable } f : \mathbb{X} \times \mathbb{Y} \to \mathbb{R},$$

where

$$h_f(x) := \mathbb{E}[f(x, Y)] \text{ for all } x \in X,$$

i.e.

$$\mathbb{E}[f(X, Y)|X] = \mathbb{E}[f(x, Y)]|_{x=X} \text{ a.s.}$$

Exercise 14.5. Suppose $(\Omega, \mathcal{B}, P)$ and $(\Omega', \mathcal{B}', P')$ are two probability spaces, $(\mathbb{X}, \mathcal{M})$ and $(\mathbb{Y}, \mathcal{N})$ are measurable spaces, $X : \Omega \to \mathbb{X}$, $X' : \Omega' \to \mathbb{X}$, $Y : \Omega \to \mathbb{Y}$,and $Y' : \Omega \to \mathbb{Y}$ are measurable functions such that $P \circ (X, Y)^{-1} = P' \circ (X', Y')$, i.e. $(X, Y) \overset{d}{=} (X', Y')$. If $f : (\mathbb{X} \times \mathbb{Y}, \mathcal{M} \otimes \mathcal{N}) \to \mathbb{R}$ is a bounded measurable function and $\tilde{f} : (\mathbb{X}, \mathcal{M}) \to \mathbb{R}$ is a measurable function such that $\tilde{f}(X) = \mathbb{E}[f(X, Y)|X]$ P - a.s. then

$$\mathbb{E}'[f(X', Y')|X'] = \tilde{f}(X') \;\; P' \text{ a.s.}$$

14.2 Additional Properties of Conditional Expectations

The next theorem is devoted to extending the notion of conditional expectations to all non-negative functions and to proving conditional versions of the MCT, DCT, and Fatou's lemma.

Theorem 14.15 (Extending $\mathbb{E}_{\mathcal{G}}$). *If* $f : \Omega \to [0, \infty]$ *is* $\mathcal{B}$ *– measurable, the function* $F := \uparrow \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n]$ *exists a.s. and is, up to sets of measure zero, uniquely determined by as the* $\mathcal{G}$ *– measurable function,* $F : \Omega \to [0, \infty]$, *satisfying*

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}. \tag{14.25}$$

Hence it is consistent to denote F *by* $\mathbb{E}_{\mathcal{G}} f$. *In addition we now have;*

1. *Properties 2., 5. (with* $0 \le g \in \mathcal{G}_b$*), and 6. of Theorem 14.4 still hold for any* $\mathcal{B}$ *– measurable functions such that* $0 \le f \le g$. *Namely;*
   a) ***Order Preserving.*** $\mathbb{E}_{\mathcal{G}} f \le \mathbb{E}_{\mathcal{G}} g$ *a.s. when* $0 \le f \le g$,

b) ***Pull out Property.*** *$\mathbb{E}_{\mathcal{G}}[hf] = h\mathbb{E}_{\mathcal{G}}[f]$ a.s. for all $h \geq 0$ and $\mathcal{G}$ – measurable.*

c) ***Tower or smoothing property.*** *If $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \mathcal{B}$. Then*

$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s.}$$

2. ***Conditional Monotone Convergence (cMCT).*** *Suppose that, almost surely, $0 \leq f_n \leq f_{n+1}$ for all n, then then $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} f_n]$ a.s.*
3. ***Conditional Fatou's Lemma (cFatou).*** *Suppose again that $0 \leq f_n \in L^1(\Omega, \mathcal{B}, P)$ a.s., then*

$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] \leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f_n] \text{ a.s.} \tag{14.26}$$

4. ***Conditional Dominated Convergence (cDCT).*** *If $f_n \to f$ a.s. and $|f_n| \leq g \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ a.s.*

Remark 14.16. Regarding item 4. above. Suppose that $f_n \overset{P}{\to} f$, $|f_n| \leq g_n \in L^1(\Omega, \mathcal{B}, P)$, $g_n \overset{P}{\to} g \in L^1(\Omega, \mathcal{B}, P)$ and $\mathbb{E}g_n \to \mathbb{E}g$. Then by the DCT in Corollary 12.8, we know that $f_n \to f$ in $L^1(\Omega, \mathcal{B}, P)$. Since $\mathbb{E}_{\mathcal{G}}$ is a contraction, it follows that $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ in $L^1(\Omega, \mathcal{B}, P)$ and hence in probability.

Proof. Since $f \wedge n \in L^1(\Omega, \mathcal{B}, P)$ and $f \wedge n$ is increasing, it follows that $F :=\uparrow \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n]$ exists a.s. Moreover, by two applications of the standard MCT, we have for any $A \in \mathcal{G}$, that

$$\mathbb{E}[F : A] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}_{\mathcal{G}}[f \wedge n] : A] = \lim_{n\to\infty} \mathbb{E}[f \wedge n : A] = \lim_{n\to\infty} \mathbb{E}[f : A].$$

Thus Eq. (14.25) holds and this uniquely determines F follows from Lemma 7.24.

Item 1. a) If $0 \leq f \leq g$, then

$$\mathbb{E}_{\mathcal{G}} f = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n] \leq \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[g \wedge n] = \mathbb{E}_{\mathcal{G}} g \text{ a.s.}$$

and so $\mathbb{E}_{\mathcal{G}}$ still preserves order. We will prove items 1b and 1c at the end of this proof.

Item 2. Suppose that, almost surely, $0 \leq f_n \leq f_{n+1}$ for all n, then $\mathbb{E}_{\mathcal{G}} f_n$ is a.s. increasing in n. Hence, again by two applications of the MCT, for any $A \in \mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}\left[\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n : A\right] &= \lim_{n\to\infty} \mathbb{E}[\mathbb{E}_{\mathcal{G}} f_n : A] = \lim_{n\to\infty} \mathbb{E}[f_n : A] \\ &= \mathbb{E}\left[\lim_{n\to\infty} f_n : A\right] = \mathbb{E}\left[\mathbb{E}_{\mathcal{G}}\left[\lim_{n\to\infty} f_n\right] : A\right]\end{aligned}$$

from which it follows that $\lim_{n\to\infty} \mathbb{E}_{\mathcal{G}} f_n = \mathbb{E}_{\mathcal{G}}[\lim_{n\to\infty} f_n]$ a.s.

Item 3. For $0 \leq f_n$, let $g_k := \inf_{n \geq k} f_n$. Then $g_k \leq f_k$ for all k and $g_k \uparrow \liminf_{n\to\infty} f_n$ and hence by cMCT and item 1.,

$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] = \lim_{k\to\infty} \mathbb{E}_{\mathcal{G}} g_k \leq \liminf_{k\to\infty} \mathbb{E}_{\mathcal{G}} f_k \text{ a.s.}$$

Item 4. As usual it suffices to consider the real case. Let $f_n \to f$ a.s. and $|f_n| \leq g$ a.s. with $g \in L^1(\Omega, \mathcal{B}, P)$. Then following the proof of the Dominated convergence theorem, we start with the fact that $0 \leq g \pm f_n$ a.s. for all n. Hence by cFatou,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(g \pm f) &= \mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty}(g \pm f_n)\right] \\ &\leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(g \pm f_n) = \mathbb{E}_{\mathcal{G}} g + \begin{cases} \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } + \text{ case} \\ -\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) & \text{in } - \text{ case,}\end{cases}\end{aligned}$$

where the above equations hold a.s. Cancelling $\mathbb{E}_{\mathcal{G}} g$ from both sides of the equation then implies

$$\limsup_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \leq \mathbb{E}_{\mathcal{G}} f \leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}(f_n) \text{ a.s.}$$

Item 1. b) If $h \geq 0$ is a $\mathcal{G}$ – measurable function and $f \geq 0$, then by cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}[hf] &\overset{\text{cMCT}}{=} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[(h \wedge n)(f \wedge n)] \\ &= \lim_{n\to\infty} (h \wedge n)\mathbb{E}_{\mathcal{G}}[(f \wedge n)] \overset{\text{cMCT}}{=} h\mathbb{E}_{\mathcal{G}} f \text{ a.s.}\end{aligned}$$

Item 1. c) Similarly by multiple uses of cMCT,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f &= \mathbb{E}_{\mathcal{G}_0} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}(f \wedge n) \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f &= \mathbb{E}_{\mathcal{G}_1} \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0}[f \wedge n] \\ &= \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}_0}(f \wedge n) = \mathbb{E}_{\mathcal{G}_0} f.\end{aligned}$$

■

The next result in Lemma 14.18 shows how to localize conditional expectations. We first need the following definition.

Definition 14.17. *Suppose that $\mathcal{F}$ and $\mathcal{G}$ are sub-σ-fields of $\mathcal{B}$ and $A \in \mathcal{B}$. We say that $\mathcal{F} = \mathcal{G}$ on A iff $A \in \mathcal{F} \cap \mathcal{G}$ and $\mathcal{F}_A = \mathcal{G}_A$. Recall that $\mathcal{F}_A = \{B \cap A : B \in \mathcal{F}\}$.*

Notice that if $\mathcal{F}=\mathcal{G}$ on A then $\mathcal{F}=\mathcal{G}=\mathcal{F}\cap\mathcal{G}$ on A as well, i.e. if $A\in\mathcal{F}\cap\mathcal{G}$ and $\mathcal{F}_A=\mathcal{G}_A$ then

$$\mathcal{F}_A=\mathcal{G}_A=[\mathcal{F}\cap\mathcal{G}]_A. \tag{14.27}$$

To prove this first observe $\mathcal{F}\cap\mathcal{G}\subset\mathcal{F}$ implies $[\mathcal{F}\cap\mathcal{G}]_A\subset\mathcal{F}_A=\mathcal{G}_A$. Conversely, $B\in\mathcal{F}_A=\mathcal{G}_A$, then $B\cap A\in\mathcal{F}\cap\mathcal{G}$, i.e. $B\in[\mathcal{F}\cap\mathcal{G}]_A$.

Lemma 14.18 (Localizing Conditional Expectations). *Let $(\Omega,\mathcal{B},P)$ be a probability space, $\mathcal{F}$ and $\mathcal{G}$ be sub-sigma-fileds of $\mathcal{B}$, $X,Y\in L^1(\Omega,\mathcal{B},P)$ or $X,Y:(\Omega,\mathcal{B})\to[0,\infty]$ are measurable, and $A\in\mathcal{F}\cap\mathcal{G}$. If $\mathcal{F}=\mathcal{G}$ on A and $X=Y$ a.s. on A, then*

$$\mathbb{E}_{\mathcal{F}}X=\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}X=\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y=\mathbb{E}_{\mathcal{G}}Y \text{ a.s. on } A. \tag{14.28}$$

Alternatively put, if $A\in\mathcal{F}\cap\mathcal{G}$ and $\mathcal{F}_A=\mathcal{G}_A$ then

$$1_A\mathbb{E}_{\mathcal{F}}=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}=1_A\mathbb{E}_{\mathcal{G}}. \tag{14.29}$$

Proof. Let us start with the observation that if X is an $\mathcal{F}$ – measurable random variable, then 1_AX is $\mathcal{F}\cap\mathcal{G}$ measurable. This can be checked either directly (see Remark 14.19 below) or as follows. If $X=1_B$ with $B\in\mathcal{F}$, then $1_A1_B=1_{A\cap B}$ and $A\cap B\in\mathcal{F}_A=\mathcal{G}_A=[\mathcal{F}\cap\mathcal{G}]_A\subset\mathcal{F}\cap\mathcal{G}$ and so 1_A1_B is $\mathcal{F}\cap\mathcal{G}$ – measurable. The general X case now follows by linearity and then passing to the limit.

Suppose $X\in L^1(\Omega,\mathcal{B},P)$ or $X\ge 0$ and let $\bar{X}$ be a representative of $\mathbb{E}_{\mathcal{F}}X$. By the previous observation, $1_A\bar{X}$ is $\mathcal{F}\cap\mathcal{G}$ – measurable. Therefore,

$$1_A\bar{X}=\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[1_A\bar{X}\right]=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[\bar{X}\right]=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}\left[\mathbb{E}_{\mathcal{F}}X\right]=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}X \text{ a.s.},$$

i.e. $1_A\mathbb{E}_{\mathcal{F}}X=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}X$ a.s. This proves the first equality in Eq. (14.29) while the second follows by interchanging the roles of $\mathcal{F}$ and $\mathcal{G}$.

Equation (14.28) is now easily verified. First notice that $X=Y$ a.s. on A iff $1_AX=1_AY$ a.s.. Now from Eq. (14.29), the tower property of conditional expectation, and the fact that $1_A=1_A\cdot 1_A$,we find

$$1_A\mathbb{E}_{\mathcal{F}}X=1_A\mathbb{E}_{\mathcal{F}}[1_AX]=1_A\mathbb{E}_{\mathcal{F}}[1_AY]=1_A\mathbb{E}_{\mathcal{F}}Y=1_A\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y$$

from which it follows that $\mathbb{E}_{\mathcal{F}}X=\mathbb{E}_{\mathcal{F}\cap\mathcal{G}}Y$ a.s. on A. ■

Remark 14.19. For the direct verification that 1_AX is $\mathcal{F}\cap\mathcal{G}$ measurable, we have,

$$\{1_AX\neq 0\}=A\cap\{X\neq 0\}\in\mathcal{F}_A=\mathcal{G}_A=(\mathcal{F}\cap\mathcal{G})_A\subset\mathcal{F}\cap\mathcal{G}.$$

So for $B\in\mathcal{B}_{\mathbb{R}}$,

$$\{1_AX\in B\}=A\cap\{X\in B\}\in\mathcal{F}_A\subset\mathcal{F}\cap\mathcal{G} \text{ if } 0\notin B$$

while if $0\in B$,

$$\begin{aligned}\{1_AX\in B\}&=\{1_AX=0\}^c\cup A\cap\{X\in(B\setminus\{0\})\}\\&=\{1_AX\neq 0\}^c\cup A\cap\{X\in(B\setminus\{0\})\}\in\mathcal{F}\cap\mathcal{G}.\end{aligned}$$

Theorem 14.20 (Conditional Jensen's inequality). *Let $(\Omega,\mathcal{B},P)$ be a probability space, $-\infty\le a<b\le\infty$, and $\varphi:(a,b)\to\mathbb{R}$ be a convex function. Assume $f\in L^1(\Omega,\mathcal{B},P;\mathbb{R})$ is a random variable satisfying, $f\in(a,b)$ a.s. and $\varphi(f)\in L^1(\Omega,\mathcal{B},P;\mathbb{R})$. Then $\varphi(\mathbb{E}_{\mathcal{G}}f)\in L^1(\Omega,\mathcal{G},P)$,*

$$\varphi(\mathbb{E}_{\mathcal{G}}f)\le\mathbb{E}_{\mathcal{G}}[\varphi(f)] \text{ a.s.} \tag{14.30}$$

and

$$\mathbb{E}[\varphi(\mathbb{E}_{\mathcal{G}}f)]\le\mathbb{E}[\varphi(f)] \tag{14.31}$$

Proof. Let $\Lambda:=\mathbb{Q}\cap(a,b)$ – a countable dense subset of (a,b). By Theorem 12.47 (also see Lemma 12.44) and Figure 12.3 when φ is C^1)

$$\varphi(y)\ge\varphi(x)+\varphi'_-(x)(y-x) \text{ for all for all } x,y\in(a,b),$$

where $\varphi'_-(x)$ is the left hand derivative of φ at x. Taking $y=f$ and then taking conditional expectations imply,

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)]\ge\mathbb{E}_{\mathcal{G}}\left[\varphi(x)+\varphi'_-(x)(f-x)\right]=\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x) \text{ a.s.} \tag{14.32}$$

Since this is true for all $x\in(a,b)$ (and hence all x in the countable set, Λ) we may conclude that

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)]\ge\sup_{x\in\Lambda}\left[\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right] \text{ a.s.}$$

By Exercise 14.1, $\mathbb{E}_{\mathcal{G}}f\in(a,b)$, and hence it follows from Corollary 12.48 that

$$\sup_{x\in\Lambda}\left[\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right]=\varphi(\mathbb{E}_{\mathcal{G}}f) \text{ a.s.}$$

Combining the last two estimates proves Eq. (14.30).

From Eqs. (14.30) and (14.32) we infer,

$$|\varphi(\mathbb{E}_{\mathcal{G}}f)|\le|\mathbb{E}_{\mathcal{G}}[\varphi(f)]|\vee\left|\varphi(x)+\varphi'_-(x)(\mathbb{E}_{\mathcal{G}}f-x)\right|\in L^1(\Omega,\mathcal{G},P)$$

and hence $\varphi(\mathbb{E}_{\mathcal{G}}f)\in L^1(\Omega,\mathcal{G},P)$. Taking expectations of Eq. (14.30) is now allowed and immediately gives Eq. (14.31). ■

Corollary 14.21. *The conditional expectation operator, $\mathbb{E}_{\mathcal{G}}$ maps $L^p(\Omega,\mathcal{B},P)$ into $L^p(\Omega,\mathcal{B},P)$ and the map remains a contraction for all $1\le p\le\infty$.*

Proof. The case $p=\infty$ and $p=1$ have already been covered in Theorem 14.4. So now suppose, $1<p<\infty$, and apply Jensen's inequality with $\varphi(x)=|x|^p$ to find $|\mathbb{E}_{\mathcal{G}}f|^p\le\mathbb{E}_{\mathcal{G}}|f|^p$ a.s. Taking expectations of this inequality gives the desired result. ■

14.3 Regular Conditional Distributions

Lemma 14.22. *Suppose that $(\mathbb{X},\mathcal{M})$ is a measurable space and $F:\mathbb{X}\times\mathbb{R}\to\mathbb{R}$ is a function such that; 1) $F(\cdot,t):\mathbb{X}\to\mathbb{R}$ is $\mathcal{M}/\mathcal{B}_\mathbb{R}$ – measurable for all $t\in\mathbb{R}$, and 2) $F(x,\cdot):\mathbb{R}\to\mathbb{R}$ is right continuous for all $x\in\mathbb{X}$. Then F is $\mathcal{M}\otimes\mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R}$ – measurable.*

Proof. For $n\in\mathbb{N}$, the function,

$$F_n(x,t):=\sum_{k=-\infty}^{\infty}F\left(x,(k+1)2^{-n}\right)1_{(k2^{-n},(k+1)2^{-n}]}(t),$$

is $\mathcal{M}\otimes\mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R}$ – measurable. Using the right continuity assumption, it follows that $F(x,t)=\lim_{n\to\infty}F_n(x,t)$ for all $(x,t)\in\mathbb{X}\times\mathbb{R}$ and therefore F is also $\mathcal{M}\otimes\mathcal{B}_\mathbb{R}/\mathcal{B}_\mathbb{R}$ – measurable. ■

Theorem 14.23. *Suppose that $(\mathbb{X},\mathcal{M})$ is a measurable space, $X:\Omega\to\mathbb{X}$ is a measurable function and $Y:\Omega\to\mathbb{R}$ is a random variable. Then there exists a probability kernel, Q, on $\mathbb{X}\times\mathbb{R}$ such that $\mathbb{E}[f(Y)|X]=Q(X,f)$, P – a.s., for all bounded measurable functions, $f:\mathbb{R}\to\mathbb{R}$.*

Proof. For each $r\in\mathbb{Q}$, let $q_r:\mathbb{X}\to[0,1]$ be a measurable function such that

$$\mathbb{E}[1_{Y\le r}|X]=q_r(X)\ \text{a.s.}$$

Let $\nu:=P\circ X^{-1}$ be the law of X. Then using the basic properties of conditional expectation, $q_r\le q_s$ ν – a.s. for all $r\le s$, $\lim_{r\uparrow\infty}q_r=1$ and $\lim_{r\downarrow\infty}q_r=0$, ν – a.s. Hence the set, $\mathbb{X}_0\subset\mathbb{X}$ where $q_r(x)\le q_s(x)$ for all $r\le s$, $\lim_{r\uparrow\infty}q_r(x)=1$, and $\lim_{r\downarrow\infty}q_r(x)=0$ satisfies, $\nu(\mathbb{X}_0)=P(X\in\mathbb{X}_0)=1$. For $t\in\mathbb{R}$, let

$$F(x,t):=1_{\mathbb{X}_0}(x)\cdot\inf\{q_r(x):r>t\}+1_{\mathbb{X}\setminus\mathbb{X}_0}(x)\cdot 1_{t\ge 0}.$$

Then $F(\cdot,t):\mathbb{X}\to\mathbb{R}$ is measurable for each $t\in\mathbb{R}$ and $F(x,\cdot)$ is a distribution function on $\mathbb{R}$ for each $x\in\mathbb{X}$. Hence an application of Lemma 14.22 shows $F:\mathbb{X}\times\mathbb{R}\to[0,1]$ is measurable.

For each $x\in\mathbb{X}$ and $B\in\mathcal{B}_\mathbb{R}$, let $Q(x,B)=\mu_{F(x,\cdot)}(B)$ where μ_F denotes the probability measure on $\mathbb{R}$ determined by a distribution function, $F:\mathbb{R}\to[0,1]$.

We will now show that Q is the desired probability kernel. To prove this, let $\mathbb{H}$ be the collection of bounded measurable functions, $f:\mathbb{R}\to\mathbb{R}$, such that $\mathbb{X}\ni x\to Q(x,f)\in\mathbb{R}$ is measurable and $\mathbb{E}[f(Y)|X]=Q(X,f)$, P – a.s. It is easily seen that $\mathbb{H}$ is a linear subspace which is closed under bounded convergence. We will finish the proof by showing that $\mathbb{H}$ contains the multiplicative class, $\mathbb{M}=\left\{1_{(-\infty,t]}:t\in\mathbb{R}\right\}$.

Notice that $Q\left(x,1_{(-\infty,t]}\right)=F(x,t)$ is measurable. Now let $r\in\mathbb{Q}$ and $g:\mathbb{X}\to\mathbb{R}$ be a bounded measurable function, then

$$\begin{aligned}\mathbb{E}[1_{Y\le r}\cdot g(X)]&=\mathbb{E}[\mathbb{E}[1_{Y\le r}|X]\,g(X)]=\mathbb{E}[q_r(X)\,g(X)]\\&=\mathbb{E}[q_r(X)\,1_{\mathbb{X}_0}(X)\,g(X)].\end{aligned}$$

For $t\in\mathbb{R}$, we may let $r\downarrow t$ in the above equality (use DCT) to learn,

$$\mathbb{E}[1_{Y\le t}\cdot g(X)]=\mathbb{E}[F(X,t)\,1_{\mathbb{X}_0}(X)\,g(X)]=\mathbb{E}[F(X,t)\,g(X)].$$

Since g was arbitrary, we may conclude that

$$Q\left(X,1_{(-\infty,t]}\right)=F(X,t)=\mathbb{E}[1_{Y\le t}|X]\ \text{a.s.}$$

This completes the proof. ■

This result leads fairly immediately to the following far reaching generalization.

Theorem 14.24. *Suppose that $(\mathbb{X},\mathcal{M})$ is a measurable space and $(\mathbb{Y},\mathcal{N})$ is a standard Borel space, see Appendix 9.10. Suppose that $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ are measurable functions. Then there exists a probability kernel, Q, on $\mathbb{X}\times\mathbb{Y}$ such that $\mathbb{E}[f(Y)|X]=Q(X,f)$, P – a.s., for all bounded measurable functions, $f:\mathbb{Y}\to\mathbb{R}$.*

Proof. By definition of a standard Borel space, we may assume that $\mathbb{Y}\in\mathcal{B}_\mathbb{R}$ and $\mathcal{N}=\mathcal{B}_\mathbb{Y}$. In this case Y may also be viewed to be a measurable map form $\Omega\to\mathbb{R}$ such that $Y(\Omega)\subset\mathbb{Y}$. By Theorem 14.23, we may find a probability kernel, Q_0, on $\mathbb{X}\times\mathbb{R}$ such that

$$\mathbb{E}[f(Y)|X]=Q_0(X,f),\ P-a.s., \tag{14.33}$$

for all bounded measurable functions, $f:\mathbb{R}\to\mathbb{R}$.

Taking $f=1_\mathbb{Y}$ in Eq. (14.33) shows

$$1=\mathbb{E}[1_\mathbb{Y}(Y)|X]=Q_0(X,\mathbb{Y})\ \text{a.s..}$$

Thus if we let $\mathbb{X}_0:=\{x\in\mathbb{X}:Q_0(x,\mathbb{Y})=1\}$, we know that $P(X\in\mathbb{X}_0)=1$. Let us now define

$$Q(x,B):=1_{\mathbb{X}_0}(x)\,Q_0(x,B)+1_{\mathbb{X}\setminus\mathbb{X}_0}(x)\,\delta_y(B)\ \text{ for }\ (x,B)\in\mathbb{X}\times\mathcal{B}_\mathbb{Y},$$

where y is an arbitrary but fixed point in Y. Then and hence Q is a probability kernel on $\mathbb{X}\times\mathbb{Y}$. Moreover if $B\in\mathcal{B}_\mathbb{Y}\subset\mathcal{B}_\mathbb{R}$, then

$$Q(X,B)=1_{\mathbb{X}_0}(X)\,Q_0(X,B)=1_{\mathbb{X}_0}(X)\,\mathbb{E}[1_B(Y)|X]=\mathbb{E}[1_B(Y)|X]\ \text{a.s.}$$

This shows that Q is the desired regular conditional probability. ■

Corollary 14.25. *Suppose $\mathcal{G}$ is a sub-σ – algebra, $(\mathbb{Y},\mathcal{N})$ is a standard Borel space, and $Y:\Omega\to\mathbb{Y}$ is a measurable function. Then there exists a probability kernel, Q, on $(\Omega,\mathcal{G})\times(\mathbb{Y},\mathcal{N})$ such that $\mathbb{E}[f(Y)|\mathcal{G}]=Q(\cdot,f)$, P - a.s. for all bounded measurable functions, $f:\mathbb{Y}\to\mathbb{R}$.*

Proof. This is a special case of Theorem 14.24 applied with $(\mathbb{X},\mathcal{M})=(\Omega,\mathcal{G})$ and $Y:\Omega\to\Omega$ being the identity map which is $\mathcal{B}/\mathcal{G}$ – measurable. ■

15

The Radon-Nikodym Theorem

Theorem 15.1 (A Baby Radon-Nikodym Theorem). *Suppose $(X, \mathcal{M})$ is a measurable space, λ and ν are two finite positive measures on $\mathcal{M}$ such that $\nu(A) \leq \lambda(A)$ for all $A \in \mathcal{M}$. Then there exists a measurable function, $\rho : X \to [0,1]$ such that $d\nu = \rho d\lambda$.*

Proof. If f is a non-negative simple function, then

$$\nu(f) = \sum_{a \geq 0} a\nu(f = a) \leq \sum_{a \geq 0} a\lambda(f = a) = \lambda(f).$$

In light of Theorem 6.39 and the MCT, this inequality continues to hold for all non-negative measurable functions. Furthermore if $f \in L^1(\lambda)$, then $\nu(|f|) \leq \lambda(|f|) < \infty$ and hence $f \in L^1(\nu)$ and

$$|\nu(f)| \leq \nu(|f|) \leq \lambda(|f|) \leq \lambda(X)^{1/2} \cdot \|f\|_{L^2(\lambda)}.$$

Therefore, $L^2(\lambda) \ni f \to \nu(f) \in \mathbb{C}$ is a continuous linear functional on $L^2(\lambda)$. By the Riesz representation Theorem 13.14, there exists a unique $\rho \in L^2(\lambda)$ such that

$$\nu(f) = \int_X f\rho d\lambda \text{ for all } f \in L^2(\lambda).$$

In particular this equation holds for all bounded measurable functions, $f : X \to \mathbb{R}$ and for such a function we have

$$\nu(f) = \operatorname{Re}\nu(f) = \operatorname{Re}\int_X f\rho d\lambda = \int_X f \operatorname{Re}\rho \, d\lambda. \tag{15.1}$$

Thus by replacing ρ by $\operatorname{Re}\rho$ if necessary we may assume ρ is real.

Taking $f = 1_{\rho<0}$ in Eq. (15.1) shows

$$0 \leq \nu(\rho < 0) = \int_X 1_{\rho<0}\rho \, d\lambda \leq 0,$$

from which we conclude that $1_{\rho<0}\rho = 0$, λ – a.e., i.e. $\lambda(\rho < 0) = 0$. Therefore $\rho \geq 0$, λ – a.e. Similarly for $\alpha > 1$,

$$\lambda(\rho > \alpha) \geq \nu(\rho > \alpha) = \int_X 1_{\rho>\alpha}\rho \, d\lambda \geq \alpha\lambda(\rho > \alpha)$$

which is possible iff $\lambda(\rho > \alpha) = 0$. Letting $\alpha \downarrow 1$, it follows that $\lambda(\rho > 1) = 0$ and hence $0 \leq \rho \leq 1$, λ - a.e. ∎

Definition 15.2. *Let μ and ν be two positive measure on a measurable space, $(X, \mathcal{M})$. Then:*

1. *μ and ν are **mutually singular** (written as $\mu \perp \nu$) if there exists $A \in \mathcal{M}$ such that $\nu(A) = 0$ and $\mu(A^c) = 0$. We say that ν lives on A and μ lives on A^c.*
2. *The measure ν **is absolutely continuous relative to** μ (written as $\nu \ll \mu$) provided $\nu(A) = 0$ whenever $\mu(A) = 0$.*

As an example, suppose that μ is a positive measure and $\rho \geq 0$ is a measurable function. Then the measure, $\nu := \rho\mu$ is absolutely continuous relative to μ. Indeed, if $\mu(A) = 0$ then

$$\nu(A) = \int_A \rho d\mu = 0.$$

We will eventually show that if μ and ν are σ – finite and $\nu \ll \mu$, then $d\nu = \rho d\mu$ for some measurable function, $\rho \geq 0$.

Definition 15.3 (Lebesgue Decomposition). *Let μ and ν be two positive measure on a measurable space, $(X, \mathcal{M})$. Two positive measures ν_a and ν_s form a **Lebesgue decomposition** of ν relative to μ if $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$, and $\nu_s \perp \mu$.*

Lemma 15.4. *If μ_1, μ_2 and ν are positive measures on $(X, \mathcal{M})$ such that $\mu_1 \perp \nu$ and $\mu_2 \perp \nu$, then $(\mu_1 + \mu_2) \perp \nu$. More generally if $\{\mu_i\}_{i=1}^\infty$ is a sequence of positive measures such that $\mu_i \perp \nu$ for all i then $\mu = \sum_{i=1}^\infty \mu_i$ is singular relative to ν.*

Proof. It suffices to prove the second assertion since we can then take $\mu_j \equiv 0$ for all $j \geq 3$. Choose $A_i \in \mathcal{M}$ such that $\nu(A_i) = 0$ and $\mu_i(A_i^c) = 0$ for all i. Letting $A := \cup_i A_i$ we have $\nu(A) = 0$. Moreover, since $A^c = \cap_i A_i^c \subset A_m^c$ for all m, we have $\mu_i(A^c) = 0$ for all i and therefore, $\mu(A^c) = 0$. This shows that $\mu \perp \nu$. ∎

Lemma 15.5. *Let ν and μ be positive measures on $(X, \mathcal{M})$. If there exists a Lebesgue decomposition, $\nu = \nu_s + \nu_a$, of the measure ν relative to μ then this decomposition is unique. Moreover: if ν is a σ – finite measure then so are ν_s and ν_a.*

Proof. Since $\nu_s \perp \mu$, there exists $A \in \mathcal{M}$ such that $\mu(A) = 0$ and $\nu_s(A^c) = 0$ and because $\nu_a \ll \mu$, we also know that $\nu_a(A) = 0$. So for $C \in \mathcal{M}$,

$$\nu(C \cap A) = \nu_s(C \cap A) + \nu_a(C \cap A) = \nu_s(C \cap A) = \nu_s(C) \tag{15.2}$$

and

$$\nu(C \cap A^c) = \nu_s(C \cap A^c) + \nu_a(C \cap A^c) = \nu_a(C \cap A^c) = \nu_a(C). \tag{15.3}$$

Now suppose we have another Lebesgue decomposition, $\nu = \tilde{\nu}_a + \tilde{\nu}_s$ with $\tilde{\nu}_s \perp \mu$ and $\tilde{\nu}_a \ll \mu$. Working as above, we may choose $\tilde{A} \in \mathcal{M}$ such that $\mu(\tilde{A}) = 0$ and $\tilde{A}^c$ is $\tilde{\nu}_s$ – null. Then $B = A \cup \tilde{A}$ is still a μ – null set and and $B^c = A^c \cap \tilde{A}^c$ is a null set for both ν_s and $\tilde{\nu}_s$. Therefore we may use Eqs. (15.2) and (15.3) with A being replaced by B to conclude,

$$\nu_s(C) = \nu(C \cap B) = \tilde{\nu}_s(C) \text{ and}$$
$$\nu_a(C) = \nu(C \cap B^c) = \tilde{\nu}_a(C) \text{ for all } C \in \mathcal{M}.$$

Lastly if ν is a σ – finite measure then there exists $X_n \in \mathcal{M}$ such that $X = \sum_{n=1}^\infty X_n$ and $\nu(X_n) < \infty$ for all n. Since $\infty > \nu(X_n) = \nu_a(X_n) + \nu_s(X_n)$, we must have $\nu_a(X_n) < \infty$ and $\nu_s(X_n) < \infty$, showing ν_a and ν_s are σ – finite as well. ■

Lemma 15.6. *Suppose μ is a positive measure on $(X, \mathcal{M})$ and $f, g : X \to [0, \infty]$ are functions such that the measures, $fd\mu$ and $gd\mu$ are σ – finite and further satisfy,*

$$\int_A f d\mu = \int_A g d\mu \text{ for all } A \in \mathcal{M}. \tag{15.4}$$

Then $f(x) = g(x)$ for μ – a.e. x.

Proof. By assumption there exists $X_n \in \mathcal{M}$ such that $X_n \uparrow X$ and $\int_{X_n} f d\mu < \infty$ and $\int_{X_n} g d\mu < \infty$ for all n. Replacing A by $A \cap X_n$ in Eq. (15.4) implies

$$\int_A 1_{X_n} f d\mu = \int_{A \cap X_n} f d\mu = \int_{A \cap X_n} g d\mu = \int_A 1_{X_n} g d\mu$$

for all $A \in \mathcal{M}$. Since $1_{X_n} f$ and $1_{X_n} g$ are in $L^1(\mu)$ for all n, this equation implies $1_{X_n} f = 1_{X_n} g$, μ – a.e. Letting $n \to \infty$ then shows that $f = g$, μ – a.e. ■

Remark 15.7. Lemma 15.6 is in general false without the σ – finiteness assumption. A trivial counterexample is to take $\mathcal{M} = 2^X$, $\mu(A) = \infty$ for all non-empty $A \in \mathcal{M}$, $f = 1_X$ and $g = 2 \cdot 1_X$. Then Eq. (15.4) holds yet $f \neq g$.

Theorem 15.8 (Radon Nikodym Theorem for Positive Measures). *Suppose that μ and ν are σ – finite positive measures on $(X, \mathcal{M})$. Then ν has a unique Lebesgue decomposition $\nu = \nu_a + \nu_s$ relative to μ and there exists a unique (modulo sets of μ – measure 0) function $\rho : X \to [0, \infty)$ such that $d\nu_a = \rho d\mu$. Moreover, $\nu_s = 0$ iff $\nu \ll \mu$.*

Proof. The uniqueness assertions follow directly from Lemmas 15.5 and 15.6.

Existence when μ and ν are both finite measures. (Von-Neumann's Proof. See Remark 15.9 for the motivation for this proof.) First suppose that μ and ν are **finite** measures and let $\lambda = \mu + \nu$. By Theorem 15.1, $d\nu = hd\lambda$ with $0 \leq h \leq 1$ and this implies, for all non-negative measurable functions f, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh) \tag{15.5}$$

or equivalently

$$\nu(f(1-h)) = \mu(fh). \tag{15.6}$$

Taking $f = 1_{\{h=1\}}$ in Eq. (15.6) shows that

$$\mu(\{h = 1\}) = \nu(1_{\{h=1\}}(1-h)) = 0,$$

i.e. $0 \leq h(x) < 1$ for μ - a.e. x. Let

$$\rho := 1_{\{h<1\}} \frac{h}{1-h}$$

and then take $f = g 1_{\{h<1\}}(1-h)^{-1}$ with $g \geq 0$ in Eq. (15.6) to learn

$$\nu(g 1_{\{h<1\}}) = \mu(g 1_{\{h<1\}}(1-h)^{-1} h) = \mu(\rho g).$$

Hence if we define

$$\nu_a := 1_{\{h<1\}} \nu \text{ and } \nu_s := 1_{\{h=1\}} \nu,$$

we then have $\nu_s \perp \mu$ (since ν_s "lives" on $\{h = 1\}$ while $\mu(h = 1) = 0$) and $\nu_a = \rho\mu$ and in particular $\nu_a \ll \mu$. Hence $\nu = \nu_a + \nu_s$ is the desired Lebesgue decomposition of ν. If we further assume that $\nu \ll \mu$, then $\mu(h = 1) = 0$ implies $\nu(h = 1) = 0$ and hence that $\nu_s = 0$ and we conclude that $\nu = \nu_a = \rho\mu$.

Existence when μ and ν are σ-finite measures. Write $X = \sum_{n=1}^\infty X_n$ where $X_n \in \mathcal{M}$ are chosen so that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$ for all n. Let $d\mu_n = 1_{X_n} d\mu$ and $d\nu_n = 1_{X_n} d\nu$. Then by what we have just proved there exists $\rho_n \in L^1(X, \mu_n) \subset L^1(X, \mu)$ and measure ν_n^s such that $d\nu_n = \rho_n d\mu_n + d\nu_n^s$ with $\nu_n^s \perp \mu_n$. Since μ_n and ν_n^s "live" on X_n there exists $A_n \in \mathcal{M}_{X_n}$ such that $\mu(A_n) = \mu_n(A_n) = 0$ and

$$\nu_n^s\left(X\setminus A_n\right)=\nu_n^s\left(X_n\setminus A_n\right)=0.$$

This shows that $\nu_n^s \perp \mu$ for all n and so by Lemma 15.4, $\nu_s := \sum_{n=1}^\infty \nu_n^s$ is singular relative to μ. Since

$$\nu=\sum_{n=1}^{\infty}\nu_n=\sum_{n=1}^{\infty}\left(\rho_n\mu_n+\nu_n^s\right)=\sum_{n=1}^{\infty}\left(\rho_n 1_{X_n}\mu+\nu_n^s\right)=\rho\mu+\nu_s, \tag{15.7}$$

where $\rho := \sum_{n=1}^\infty 1_{X_n}\rho_n$, it follows that $\nu=\nu_a+\nu_s$ with $\nu_a=\rho\mu$. Hence this is the desired Lebesgue decomposition of ν relative to μ. ■

Remark 15.9. Here is the motivation for the above construction. Suppose that $d\nu = d\nu_s+\rho d\mu$ is the Radon-Nikodym decomposition and $X=A\sum B$ such that $\nu_s(B)=0$ and $\mu(A)=0$. Then we find

$$\nu_s(f)+\mu(\rho f)=\nu(f)=\lambda(hf)=\nu(hf)+\mu(hf).$$

Letting $f\to 1_A f$ then implies that

$$\nu(1_Af)=\nu_s(1_Af)=\nu(1_Ahf)$$

which show that $h=1$, ν –a.e. on A. Also letting $f\to 1_Bf$ implies that

$$\mu(\rho 1_Bf)=\nu(h1_Bf)+\mu(h1_Bf)=\mu(\rho h1_Bf)+\mu(h1_Bf)$$

which implies, $\rho=\rho h+h$, μ – a.e. on B, i.e.

$$\rho\left(1-h\right)=h,\ \mu\text{– a.e. on } B.$$

In particular it follows that $h<1$, $\mu=\nu$ – a.e. on B and that $\rho=\frac{h}{1-h}1_{h<1}$, μ – a.e. So up to sets of ν – measure zero, $A=\{h=1\}$ and $B=\{h<1\}$ and therefore,

$$d\nu=1_{\{h=1\}}d\nu+1_{\{h<1\}}d\nu=1_{\{h=1\}}d\nu+\frac{h}{1-h}1_{h<1}d\mu.$$

Some Ergodic Theory

Theorem 16.1 (Von-Neumann's Mean Ergodic Theorem). *(For more on Ergodic Theory, see [4] and [2].) Let $U : H \to H$ be an isometry on a Hilbert space H, $M = \mathrm{Nul}(U-I)$, $P = P_M$ be orthogonal projection onto M, and $S_n = \frac{1}{n}\sum_{k=0}^{n-1} U^k$. Show $S_n \to P_M$* ***strongly*** *by which we mean $\lim_{n\to\infty} S_n x = P_M x$ for all $x \in H$.*

Proof. Since U is an isometry we have $(Ux, Uy) = (x, y)$ for all $x, y \in H$ and therefore that $U^*U = I$. In general it is not true that $UU^* = I$ – this only happens if U is surjective, i.e. U is unitary. In general we will have $UU^* = P_{\mathrm{Ran}(U)}$.

Before starting the proof in earnest we need to prove $\mathrm{Nul}(U^*-I) = \mathrm{Nul}(U-I)$. If $x \in \mathrm{Nul}(U-I)$ then $x = Ux$ and therefore $U^*x = U^*Ux = x$, i.e. $x \in \mathrm{Nul}(U^*-I)$. Conversely if $x \in \mathrm{Nul}(U^*-I)$ then $U^*x = x$ and we have

$$\|Ux - x\|^2 = 2\|x\|^2 - 2\,\mathrm{Re}\,(Ux, x) = 2\|x\|^2 - 2\,\mathrm{Re}\,(x, U^*x) = 2\|x\|^2 - 2\,\mathrm{Re}\,(x, x) = 0$$

which shows that $Ux = x$, i.e. $x \in \mathrm{Nul}\,(U-I)$. With this remark in hand we can easily complete the proof.

If $x \in M := \mathrm{Nul}\,(U-I)$, then $S_n x = x$ for all $x \in M$ and so trivially, $S_n x \to x$ for all $x \in M$. We now need to show that $S_n x \to 0$ for all

$$x \in M^\perp = \mathrm{Nul}\,(U-I)^\perp = \mathrm{Nul}\,(U^*-I)^\perp = \overline{\mathrm{Ran}(U-I)}.$$

We start with $x \in \mathrm{Ran}(U-I)$ so that $x = Uy - y$ for some $y \in H$. By a telescoping series argument we find that

$$S_n x = \frac{1}{n}\left(U^n y - y\right) \to 0 \text{ as } n \to \infty.$$

Finally if $x \in M^\perp = \overline{\mathrm{Ran}(U-I)}$ and $y \in \mathrm{Ran}(U-I)$, we have, since $\|S_n\| \le 1$, that

$$\|S_n x - S_n y\| \le \|x - y\|$$

and hence

$$\limsup_{n\to\infty} \|S_n x - S_n y\| \le \|x - y\|.$$

Therefore it follows that $\limsup_{n\to\infty}\|S_n x\| \le \|x-y\|$. Letting $y \to x$ shows that $\limsup_{n\to\infty}\|S_n x\| = 0$ for all $x \in M^\perp$. Therefore if $x \in H$ and $x = m + m^\perp \in M \oplus M^\perp$, then

$$\lim_{n\to\infty} S_n x = \lim_{n\to\infty} S_n m + \lim_{n\to\infty} S_n m^\perp = m + 0 = P_M x$$

as was to be proved. ■

For the rest of this section, suppose that $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_*\mu = \mu$. For more results along the lines of this chapter, the reader is referred to Kallenberg [6, Chapter 10]. The reader may also benefit from Norris's notes in [9].

Definition 16.2. *Let*

$$\mathcal{B}_\theta := \left\{A \in \mathcal{B} : \theta^{-1}(A) = A\right\} \text{ and}$$
$$\mathcal{B}'_\theta := \left\{A \in \mathcal{B} : \mu\left(\theta^{-1}(A)\,\Delta A\right) = 0\right\}$$

be the ***invariant*** σ *–* ***field*** *and* ***almost invariant*** σ *–fields respectively.*

Lemma 16.3. *The elements of $\mathcal{B}'_\theta$ are the same as the elements in $\mathcal{B}_\theta$ modulo null sets, i.e.*

$$\mathcal{B}'_\theta = \{B \in \mathcal{B} : \exists\ A \in \mathcal{B}_\theta \ni \mu(A \Delta B) = 0\}.$$

Proof. If $A \in \mathcal{B}_\theta$ and $B \in \mathcal{B}$ such that $\mu(A\Delta B) = 0$, then

$$\mu\left(A \Delta \theta^{-1}(B)\right) = \mu\left(\theta^{-1}(A)\,\Delta\theta^{-1}(B)\right) = \mu\theta^{-1}(A\Delta B) = \mu(A\Delta B) = 0$$

and therefore it follows that

$$\mu\left(B\Delta\theta^{-1}(B)\right) \le \mu(B\Delta A) + \mu\left(A\Delta\theta^{-1}(B)\right) = 0.$$

This shows that $B \in \mathcal{B}'_\theta$.

Conversely if $B \in \mathcal{B}'_\theta$, then by the invariance of μ under θ it follows that $\mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0$ for all $k = 0, 1, 2, 3 \ldots$. In particular we learn that

$$\begin{aligned}\mu\left(\theta^{-k}(B)\,\Delta B\right) &= \mu\left(\left|1_{\theta^{-k}(B)} - 1_B\right|\right) \\ &\le \sum_{l=0}^{k-1}\mu\left(\left|1_{\theta^{-l}(B)} - 1_{\theta^{-(l+1)}(B)}\right|\right) \\ &= \sum_{l=0}^{k-1}\mu\left(\theta^{-l}(B)\,\Delta\theta^{-(l+1)}(B)\right) = 0.\end{aligned}$$

Moreover from Exercise 5.2 it follows, for each $n \in \mathbb{N}$, that

$$\mu\left(B \Delta \cup_{k \geq n} \theta^{-k}(B)\right) \leq \sum_{k=n}^{\infty} \mu\left(B \Delta \theta^{-k}(B)\right) = 0.$$

Similarly using

$$\mu([\cap B_i] \Delta [\cap A_i]) = \mu([\cup B_i^c] \Delta [\cup A_i^c]) \leq \sum_{i=1}^{\infty} \mu(B_i^c \Delta A_i^c) = \sum_{i=1}^{\infty} \mu(B_i \Delta A_i),$$

it now follows that $\mu\left(B \Delta \cap_{n=1}^{\infty} \cup_{k \geq n} \theta^{-k}(B)\right) = 0$. So if we let

$$A := \cap_{n=1}^{\infty} \cup_{k \geq n} \theta^{-k}(B) = \left\{\omega \in \Omega : \omega \in \theta^{-k}(B) \text{ i.o.}\right\} = \left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ i.o.}\right\},$$

then $\theta^{-1}(A) = A$ and we have shown $\mu(B \Delta A) = 0$. (We could have just as well taken A to be equal to $\left\{\omega \in \Omega : \theta^k(\omega) \in B \text{ a.a.}\right\}$.) ■

Lemma 16.4 (BRUCE: this is already done in Exericse 12.4). *A $\mathcal{B}$ – measurable function, $f : \Omega \to \mathbb{R}$ is (almost) invariant iff f is $\mathcal{B}_\theta$ ($\mathcal{B}'_\theta$) measurable. Moreover, if f is almost invariant, then there exists and invariant function, $g : \Omega \to \mathbb{R}$, such that $f = g$, μ – a.e.*

Proof. If f is invariant, $f \circ \theta = f$, then $\theta^{-1}(\{f \leq x\}) = \{f \circ \theta \leq x\} = \{f \leq x\}$ which shows that $\{f \leq x\} \in \mathcal{B}_\theta$ for all $x \in \mathbb{R}$ and therefore f is $\mathcal{B}_\theta$ – measurable. Similarly if $f \circ \theta = f$ μ – a.e., then

$$\mu\left(\left|1_{\theta^{-1}(\{f \leq x\})} - 1_{\{f \leq x\}}\right|\right) = \mu\left(\left|1_{\{f \circ \theta \leq x\}} - 1_{\{f \leq x\}}\right|\right) = \mu\left(\left|1_{(-\infty, x]} \circ f \circ \theta - 1_{(-\infty, x]} \circ f\right|\right) = 0$$

from which it follows that $\{f \leq x\} \in \mathcal{B}'_\theta$ for all $x \in \mathbb{R}$, that is f is $\mathcal{B}'_\theta$ – measurable.

Conversely if $f : \Omega \to \mathbb{R}$ is ($\mathcal{B}'_\theta$) $\mathcal{B}_\theta$ -measurable, then for all $-\infty < a < b < \infty$, ($\{a < f \leq b\} \in \mathcal{B}'_\theta$) $\{a < f \leq b\} \in \mathcal{B}_\theta$ from which it follows that $1_{\{a < f \leq b\}}$ is (almost) invariant. Thus for every $N \in \mathbb{N}$ the function defined by;

$$f_N := \sum_{n=-N^2}^{N^2} \frac{n}{N} 1_{\left\{\frac{n-1}{N} < f \leq \frac{n}{N}\right\}},$$

is (almost) invariant. As $f = \lim_{N \to \infty} f_N$, it follows that f is (almost) invariant as well.

In the case where f is almost invariant, we can choose $D_N(n) \in \mathcal{B}_\theta$ such that $\mu\left(D_N(n) \Delta \left\{\frac{n-1}{N} < f \leq \frac{n}{N}\right\}\right) = 0$ for all n and N and then set

$$g_N := \sum_{n=-N^2}^{N^2} \frac{n}{N} 1_{D_N(n)}.$$

We then have $g_N = f_N$ a.e. and g_N is $\mathcal{B}_\theta$ – measurable. We may thus conclude that $\tilde{g} := \limsup_{N \to \infty} g_N$ is $\mathcal{B}_\theta$ – measurable. It now follows that $g := \tilde{g} 1_{|\tilde{g}| < \infty}$ is $\mathcal{B}_\theta$ – measurable function such that $g = f$ a.e. ■

Theorem 16.5. *Suppose that $(\Omega, \mathcal{B}, \mu)$ is a σ – finite measure space and $\theta : \Omega \to \Omega$ is a measurable map such that $\theta_* \mu = \mu$. Then;*

1. *$U : L^2(\mu) \to L^2(\mu)$ defined by $Uf := f \circ \theta$ is an isometry. The isometry U is unitary if θ^{-1} exists as a measurable map.*
2. *The map,*
$$L^2(\Omega, \mathcal{B}_\theta, \mu) \ni f \to f \in \operatorname{Nul}(U - I)$$
is unitary. In other words, $Uf = f$ iff there exists $g \in L^2(\Omega, \mathcal{B}_\theta, \mu)$ such that $f = g$ a.e.
3. *For every $f \in L^2(\mu)$ we have,*
$$L^2(\mu) - \lim_{n \to \infty} \frac{f + f \circ \theta + \cdots + f \circ \theta^{n-1}}{n} = \mathbb{E}_{\mathcal{B}_\theta}[f]$$
where $\mathbb{E}_{\mathcal{B}_\theta}$ denotes orthogonal projection from $L^2(\Omega, \mathcal{B}, \mu)$ onto $L^2(\Omega, \mathcal{B}_\theta, \mu)$, i.e. $\mathbb{E}_{\mathcal{B}_\theta}$ is conditional expectation.

Proof. 1. To see that U is an isometry observe that

$$\|Uf\|^2 = \int_\Omega |f \circ \theta|^2 d\mu = \int_\Omega |f|^2 d(\theta_* \mu) = \int_\Omega |f|^2 d\mu = \|f\|^2$$

for all $f \in L^2(\mu)$.

2. $f \in \operatorname{Nul}(U - I)$ iff $f \circ \theta = Uf = f$ a.e., i.e. iff f is almost invariant. According to Lemma 16.4 this happen iff there exists a $\mathcal{B}_\theta$ – measurable function, g, such that $f = g$ a.e. Necessarily, $g \in L^2(\mu)$ so that $g \in L^2(\Omega, \mathcal{B}_\theta, \mu)$ as required.

3. The last assertion now follows from items 1. and 2. and the mean ergodic Theorem 16.1. ■

Lemma 16.6 (Maximal ergodic lemma). *Suppose $\{\xi_k\}_{k=1}^\infty$ is a stationary sequence (i.e. $(\xi_1, \xi_2, \dots) \overset{d}{=} (\xi_2, \xi_3, , \dots)$) and $S_n = \xi_1 + \cdots + \xi_n$. then*

$$\mathbb{E}\left[\xi_1 : \sup_n S_n > 0\right] \geq 0. \tag{16.1}$$

Proof. With out loss of generality, we may assume that $\Omega = \mathbb{R}^{\mathbb{N}}$, $\mathcal{B} = \mathcal{B}_{\mathbb{R}}^{\otimes \mathbb{N}}$, and $\mu = \mathrm{Law}_P(\xi_1, \xi_2, \dots)$. The stationary assumption then amounts to $\mu \circ \theta^{-1} = \mu$ where $\theta : \Omega \to \Omega$ is the shift operation, $\theta(\omega_1, \omega_2, \omega_3, \dots) = (\omega_2, \omega_3, \omega_4, \dots)$. Let $S_n^* := \max(S_1, S_2, \dots, S_n)$ for all $n \in \mathbb{N}$. We then have for $1 \le k \le n$ that

$$S_k = \xi_1 + S_{k-1} \circ \theta \le \xi_1 + S_{k-1}^* \circ \theta \le \xi_1 + S_n^* \circ \theta = \xi_1 + [S_n^* \circ \theta]_+$$

and therefore, $S_n^* \le \xi_1 + [S_n^* \circ \theta]_+$. So we may conclude that

$$\begin{aligned}\mathbb{E}[\xi_1 : S_n^* > 0] &\ge \mathbb{E}\left[S_n^* - [S_n^* \circ \theta]_+ : S_n^* > 0\right] \\ &= \mathbb{E}\left[[S_n^*]_+ - [S_n^* \circ \theta]_+ 1_{S_n^* > 0}\right] \\ &\ge \mathbb{E}\left[[S_n^*]_+ - [S_n^* \circ \theta]_+\right] = \mathbb{E}[S_n^*]_+ - \mathbb{E}[S_n^* \circ \theta]_+ = 0.\end{aligned}$$

Letting $n \to \infty$ making use of the MCT the gives Eq. (16.1). ■

Theorem 16.7 (Birkoff's Ergodic Theorem). *Suppose that $f \in L^1(\Omega, \mathcal{B}, P)$ or $f \ge 0$ and $\mathcal{B}$ – measurable, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n f \circ \theta^{k-1} = \mathbb{E}[f | \mathcal{B}_\theta] \ \text{ a.s..} \tag{16.2}$$

*Moreover if $f \in L^p(\Omega, \mathcal{B}, P)$ for some $1 \le p < \infty$ then the convergence in Eq. (**??**) holds in L^p as well.*

Proof. Let $\eta_k := f \circ \theta^{k-1}$ and $\xi_k := \eta_k - g$ where $g := \mathbb{E}[\eta_1 | \mathcal{B}_\theta] = \mathbb{E}[f | \mathcal{B}_\theta]$. Since P is θ – invariant we have

$$\begin{aligned}(\xi_1, \xi_2, \dots) &\overset{d}{=} (\xi_1, \xi_2, \dots) \circ \theta = (\eta_1 \circ \theta - g \circ \theta, \eta_2 \circ \theta - g \circ \theta, \dots) \\ &= (\eta_2 - g, \eta_3 - g, \dots) = (\xi_2, \xi_3, \dots).\end{aligned}$$

Thus ξ_i is still a stationary sequence. As above let $S_n = \xi_1 + \dots + \xi_n$ and for $\varepsilon > 0$ let

$$A_\varepsilon := \left\{\limsup_{n \to \infty} \frac{1}{n} S_n > \varepsilon\right\}, \ \xi_n^\varepsilon := (\xi_n - \varepsilon) 1_{A_\varepsilon},$$
$$\text{and } S_n^\varepsilon := \xi_1^\varepsilon + \dots + \xi_n^\varepsilon = (S_n - n\varepsilon) 1_{A_\varepsilon}.$$

Observe that

$$\left\{\sup_n S_n^\varepsilon > 0\right\} = \left\{\sup_n \frac{S_n^\varepsilon}{n} > 0\right\} = \left\{\sup_n \frac{S_n}{n} > \varepsilon\right\} \cap A_\varepsilon = A_\varepsilon$$

and so by the Maximal ergodic Lemma 16.6, we learn that

$$0 \le \mathbb{E}\left[\xi_1^\varepsilon : \sup_n S_n^\varepsilon > 0\right] = \mathbb{E}[(\xi_1 - \varepsilon) : A_\varepsilon] = \mathbb{E}[\xi_1 : A_\varepsilon] - \varepsilon P(A_\varepsilon).$$

Since $A_\varepsilon \in \mathcal{B}_\theta$ (because $\limsup_{n\to\infty} \frac{1}{n} S_n = \limsup_{n \to \infty} \frac{1}{n} S_n \circ \theta$), it follows that

$$\mathbb{E}[\xi_1 : A_\varepsilon] = \mathbb{E}[\mathbb{E}[\xi_1 | \mathcal{B}_\theta] : A_\varepsilon] = 0.$$

Thus we may conclude that $\varepsilon P(A_\varepsilon) \le 0$, i.e. $P(A_\varepsilon) = 0$. Since $\varepsilon > 0$ was arbitrary, it follows that $\limsup_{n\to\infty} \frac{1}{n} S_n \le 0$ a.s. Applying the same logic with ξ_i replaced by $-\xi_i$ for all i also shows that $\limsup_{n \to \infty} \left(-\frac{1}{n} S_n\right) \le 0$ a.s., that is $\liminf_{n\to\infty} \frac{1}{n} S_n \ge 0$ from which it follows that $\lim_{n \to \infty} \frac{1}{n} S_n = 0$ a.s.

Now suppose that $f \in L^p$ and $A \in \mathcal{B}$. Making use of Jensen's inequality relative to normalized counting measure on $\{1, 2, \dots, n\}$ we have

$$\begin{aligned}\mathbb{E}\left[1_A \left|\frac{1}{n} \sum_{k=1}^n f \circ \theta^{k-1}\right|^p\right] &\le \mathbb{E}\left[1_A \frac{1}{n} \sum_{k=1}^n \left|f \circ \theta^{k-1}\right|^p\right] \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}\left[\left|f \circ \theta^{k-1}\right|^p : A\right].\end{aligned}$$

For any $r > 0$ we also have,

$$\begin{aligned}\mathbb{E}\left[\left|f \circ \theta^{k-1}\right|^p : A\right] =& \mathbb{E}\left[\left|f \circ \theta^{k-1}\right|^p : \left|f \circ \theta^{k-1}\right| > r, A\right] \\ &+ \mathbb{E}\left[\left|f \circ \theta^{k-1}\right|^p : \left|f \circ \theta^{k-1}\right| \le r, A\right] \\ \le& \mathbb{E}\left[\left|f \circ \theta^{k-1}\right|^p : \left|f \circ \theta^{k-1}\right| > r\right] + r^p P(A) \\ =& \mathbb{E}[|f|^p : |f| > r] + r^p P(A).\end{aligned}$$

Combining the previous two estimates gives,

$$\begin{aligned}\limsup_{P(A) \to 0} \sup_n \mathbb{E}\left[1_A \left|\frac{1}{n} \sum_{k=1}^n f \circ \theta^{k-1}\right|^p\right] &\le \limsup_{P(A) \to 0} [\mathbb{E}[|f|^p : |f| > r] + r^p P(A)] \\ &= \mathbb{E}[|f|^p : |f| > r] \to 0 \text{ as } r \to \infty\end{aligned}$$

and therefore $\left\{\left|\frac{1}{n} \sum_{k=1}^n f \circ \theta^{k-1}\right|^p\right\}_{n=1}^\infty$ is uniformly integrable. The stated L^p – convergence is now a consequence of Corollary 12.39.

Finally we need to consider the case where $f \ge 0$ but not integrable. As before, let $g = \mathbb{E}[f | \mathcal{B}_\theta] \ge 0$. Let $r \in (0, \infty)$ and let $f_r := f \cdot 1_{g \le r}$. We then have

$$\mathbb{E}[f_r | \mathcal{B}_\theta] = \mathbb{E}[f \cdot 1_{g \le r} | \mathcal{B}_\theta] = 1_{g \le r} \mathbb{E}[f \cdot | \mathcal{B}_\theta] = 1_{g \le r} \cdot g$$

and in particular, $\mathbb{E} f_r = \mathbb{E}(1_{g \le r} g) \le r < \infty$. Thus by the L^1 – case already proved,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f_r\circ\theta^{k-1} = 1_{g\le r}\cdot g \text{ a.s.}$$

On the other hand, since g is θ –invariant, we see that $f_r\circ\theta^k = f\circ\theta^k\cdot 1_{g\le r}$ and therefore that

$$\frac{1}{n}\sum_{k=1}^{n} f_r\circ\theta^{k-1} = \left(\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1}\right)1_{g\le r}.$$

Using these identities and the fact that $r<\infty$ was arbitrary we may conclude that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} = g \text{ a.s. on } \{g<\infty\}.$$

To take care of the set where $\{g=\infty\}$, ageing let $r\in(0,\infty)$ and observe that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} \ge \liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\left[f\circ\theta^{k-1}\right]\wedge r = \mathbb{E}\left[f\wedge r|\mathcal{B}_\theta\right].$$

Letting $r\uparrow\infty$ and using the cMCT implies,

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} \ge \mathbb{E}\left[f|\mathcal{B}_\theta\right] = g$$

and therefore $\liminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} = \infty$ a.s. on $\{g=\infty\}$. This then shows that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\circ\theta^{k-1} = \infty = g \text{ a.s. on } \{g=\infty\}.$$

■

As a corollary we have the following version of the strong law of large numbers, also see Theorems **??** and Example **??** below for other proofs.

Theorem 16.8. *Suppose that* $\{X_n\}_{n=1}^\infty$ *are i.i.d. random variables and let* $S_n := X_1+\dots+X_n$. *If* X_n *are integrable or* $X_n\ge 0$, *then*

$$\lim_{n\to\infty}\frac{1}{n}S_n = \mathbb{E}X_1 \textit{ a.s.}$$

Proof. We may assume that $\Omega=\mathbb{R}^{\mathbb{N}}$, $\mathcal{B}$ is the product $\sigma-$algebra, and $P=\mu^{\otimes\mathbb{N}}$ where $\mu=\text{Law}(X_1)$. In this model, $X_n(\omega)=\omega_n$ for all $\omega\in\Omega$. As usual let $\theta:\Omega\to\Omega$ be the natural shift operation on Ω. Then $X_n = X_1\circ\theta^{n-1}$ and therefore, $S_n=\sum_{k=1}^n X_1\circ\theta^{k-1}$. So by Birkoff's ergodic theorem $\lim_{n\to\infty}\frac{1}{n}S_n = \mathbb{E}\left[X_1|\mathcal{B}_\theta\right] =: g$ a.s. If $A\in\mathcal{B}_\theta$, then $A=\theta^{-n}(A)\in\sigma(X_{n+1},X_{n+2},\dots)$ and therefore $A\in\mathcal{T}=\cap_n\sigma(X_{n+1},X_{n+2},\dots)$ – the tail $\sigma-$algebra. However by Kolmogorov's 0 - 1 law (Proposition 10.46), we know that $\mathcal{T}$ is almost trivial and therefore so is $\mathcal{B}_\theta$. Hence we may conclude that $g=c$ a.s. where $c\in[0,\infty]$ is a constant, see Lemma 10.45. If $X_1\ge 0$ a.s. and $\mathbb{E}X_1=\infty$ then we know that $\mathbb{E}\left[X_1|\mathcal{B}_\theta\right]=\infty$ on a set of positive measure and therefore $c=\infty$ in this case. When X_1 is integrable, the convergence in Birkoff's ergodic theorem is also in L^1 and therefore we may conclude that

$$c=\mathbb{E}c=\lim_{n\to\infty}\mathbb{E}\left[\frac{1}{n}S_n\right]=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[S_n\right]=\mathbb{E}X_1$$

and in all case we have shown $\lim_{n\to\infty}\frac{1}{n}S_n=\mathbb{E}\left[X_1|\mathcal{B}_\theta\right]=\mathbb{E}X_1$ a.s. ■

References

1. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
2. 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)
3. 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)
4. Paul R. Halmos, *Lectures on ergodic theory*, (1960), vii+101. MR MR0111817 (22 #2677)
5. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
6. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
7. Oliver Knill, *Probability and stochastic processes with applications*, Havard Web-Based, 1994.
8. 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)
9. ______, *Probability and measure*, Tech. report, Mathematics Department, University of Cambridge, 2009.
10. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
11. 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
12. H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR MR1013117 (90g:00004)
13. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)