Bruce K. Driver

Math 280 (Probability Theory) Lecture Notes

April 4, 2007 *File:prob.tex*

Contents

Part

Homework Problems:

-2

Math 280C Homework Problems

-2.1 Homework #1 (Due Friday, April 14, 2007)

- **Look at** the following Exercises from the Lecture Notes: 19.4, 19.10,
- **Look at** the following Exercises from Resnick Chapter 10: 19, 22-24,
- **Hand in** the following Exercises from the Lecture Notes: , 19.6, 19.7, 19.8, 19.9.

-1

Math 280B Homework Problems

-1.1 Homework 1. Due Monday, January 22, 2007

- Hand in from p. 114 : 4.27
- Hand in from p. 196 : 6.5, 6.7
- Hand in from p. 234–246: 7.12, 7.16, 7.33, 7.36 (assume each X_n is integrable!), 7.42

  Hints and comments.

1. For 6.7, observe that $X_n \stackrel{d}{=} \sigma_n N(0,1)$.
2. For 7.12, let $\{U_n : n = 0, 1, 2, ...\}$ be i.i.d. random variables uniformly distributed on (0,1) and take $X_0 = U_0$ and then define X_n inductively so that $X_{n+1} = X_n \cdot U_{n+1}$.
3. For 7.36; use the assumptions to bound $\mathbb{E}[X_n]$ in terms of $\mathbb{E}[X_n : X_n \le x]$. Then use the two series theorem.

-1.2 Homework 2. Due Monday, January 29, 2007

- Resnick Chapter 7: **Hand in** 7.9, 7.13.
- Resnick Chapter 7: **look at** 7.28. (For 28b, assume $\mathbb{E}[X_i X_j] \le \rho(i-j)$ for $i \ge j$. Also you may find it easier to show $\frac{S_n}{n} \to 0$ in L^2 rather than the weaker notion of in probability.)
- **Hand in** Exercise 13.2 from these notes.
- Resnick Chapter 8: **Hand in** 8.4a-d, 8.13 (Assume $\operatorname{Var}(N_n) > 0$ for all n.)

-1.3 Homework #3 Due Monday, February 5, 2007

- Resnick Chapter 8: **Look at:** 8.14, 8.20, 8.36
- Resnick Chapter 8: **Hand in** 8.7, 8.17, 8.31, 8.30* (Due 8.31 first), 8.34 *Ignore the part of the question referring to the moment generating function. **Hint:** use problem 8.31 and the convergence of types theorem.
- Also hand in Exercise 13.3 from these notes.

-1.4 Homework #4 Due Friday, February 16, 2007

- Resnick Chapter 9: **Look at:** 9.22, 9.33
- Resnick Chapter 9: **Hand in** 9.5, 9.6, 9.9 a-e., 9.10
- Also hand in Exercise from these notes: 14.2, 14.3, and 14.4.

-1.5 Homework #5 Due Friday, February 23, 2007

- Resnick Chapter 9: **Look at:** 8
- Resnick Chapter 9: **Hand in** 11, 28, 34 (assume $\sum_n \sigma_n^2 > 0$), 35 (hint: show $P[\xi_n \neq 0 \text{ i.o. }] = 0$.), 38 (Hint: make use Proposition 7.25.)

-1.6 Homework #6 Due Monday, March 5, 2007

- Look at Resnick Chapter 10: 11
- **Hand in** the following Exercises from the Lecture Notes: 12.1, 18.1, 18.2, 18.3, 18.4
- Resnick Chapter 10: **Hand in** $2^\dagger$, 5*, 7, 8**

†In part 2b, please explain what convention you are using when the denominator is 0.

*A Poisson process, $\{N(t)\}_{t\geq 0}$, with parameter λ satisfies (by definition): (i) N has *independent increments*, so that $N(s)$ and $N(t) - N(s)$ are independent; (ii) if $0 \leq u < v$ then $N(v) - N(u)$ has the Poisson distribution with parameter $\lambda(v-u)$.

****Hint:** use Exercise 12.1.

-1.7 Homework #7 Due Monday, March 12, 2007

- **Hand in** the following Exercises from the Lecture Notes: 18.5, 19.1, 19.2,
- **Hand in** Resnick Chapter 10: 14 (take $\mathcal{B}_n := \sigma(Y_0, Y_1, \ldots, Y_n)$ for the filtration), 16

-1.8 Homework #8 Due Wednesday, March 21, 2007 by 11:00AM!

- **Look at** the following Exercise from the Lecture Notes: 19.5.
- **Hand in** the following Exercises from the Lecture Notes: 19.3.
- Resnick Chapter 10: **Hand in** 15, 28, and 33.
  For #28, let $\mathcal{B}_n := \sigma(Y_1, \ldots, Y_n)$ define the filtration. **Hint:** for part b consider, $\ln X_n$.

Math 280A Homework Problems

Unless otherwise noted, all problems are from Resnick, S. A Probability Path, Birkhauser, 1999.

0.1 Homework 1. Due Friday, September 29, 2006

- p. 20-27: Look at: 9, 12, ,19, 27, 30, 36
- p. 20-27: Hand in: 5, 17, 18, 23, 40, 41

0.2 Homework 2. Due Friday, October 6, 2006

- p. 63-70: Look at: 18
- p. 63-70: Hand in: 3, 6, 7, 11, 13 and the following problem.

Exercise 0.1 (280A-2.1). Referring to the setup in Problem 7 on p. 64 of Resnick, compute the expected number of different coupons collected after buying n boxes of cereal.

0.3 Homework 3. Due Friday, October 13, 2006

- Look at from p. 63-70: 5, 14, 19
- Look at lecture notes: exercise 4.4 and read Section 5.5
- Hand in from p. 63-70: 16
- Hand in lecture note exercises: 4.1 – 4.3, 5.1 and 5.2.

0.4 Homework 4. Due Friday, October 20, 2006

- Look at from p. 85–90: 3, 7, 12, 17, 21
- Hand in from p. 85–90: 4, 6, 8, 9, 15
- Also hand in the following exercise.

Exercise 0.2 (280A-4.1). Suppose $\{f_n\}_{n=1}^{\infty}$ is a sequence of Random Variables on some measurable space. Let B be the set of ω such that $f_n(\omega)$ is convergent as $n \to \infty$. Show the set B is measurable, i.e. B is in the $\sigma-$algebra.

0.5 Homework 5. Due Friday, October 27, 2006

- Look at from p. 110–116: 3, 5
- Hand in from p. 110–116: 1, 6, 8, 18, 19

0.6 Homework 6. Due Friday, November 3, 2006

- Look at from p. 110–116: 3, 5, 28, 29
- Look at from p. 155–166: 6, 34
- Hand in from p. 110–116: 9, 11, 15, 25
- Hand in from p. 155–166: 7
- Hand in lecture note exercise: 7.1.

0.7 Homework 7. Due Monday, November 13, 2006

- Look at from p. 155–166: 13, 16, 37
- Hand in from p. 155–166: 11, 21, 26
- Hand in lecture note exercises: 8.1, 8.2, 8.19, 8.20.

0.7.1 Corrections and comments on Homework 7 (280A)

Problem 21 in Section 5.10 of Resnick should read,

$$\frac{d}{ds}P(s)=\sum_{k=1}^{\infty}kp_k s^{k-1} \text{ for } s\in[0,1].$$

Note that $P(s)=\sum_{k=0}^{\infty}p_k s^k$ is well defined and continuous (by DCT) for $s\in[-1,1]$. So the derivative makes sense to compute for $s\in(-1,1)$ with no qualifications. When $s=1$ you should interpret the derivative as the one sided derivative

$$\frac{d}{ds}|_1 P(s):=\lim_{h\downarrow 0}\frac{P(1)-P(1-h)}{h}$$

and you will need to allow for this limit to be infinite in case $\sum_{k=1}^{\infty}kp_k=\infty$. In computing $\frac{d}{ds}|_1 P(s)$, you may wish to use the fact (draw a picture or give a calculus proof) that

$$\frac{1-s^k}{1-s} \text{ increases to } k \text{ as } s\uparrow 1.$$

Hint for Exercise 8.20: Start by observing that

$$\mathbb{E}\left(\frac{S_n}{n}-\mu\right)^4 d\mu = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)\right)^4$$
$$= \frac{1}{n^4}\sum_{k,j,l,p=1}^{n}\mathbb{E}\left[(X_k-\mu)(X_j-\mu)(X_l-\mu)(X_p-\mu)\right].$$

Then analyze for which groups of indices (k, j, l, p);

$$\mathbb{E}\left[(X_k-\mu)(X_j-\mu)(X_l-\mu)(X_p-\mu)\right] \neq 0.$$

0.8 Homework 8. Due Monday, November 27, 2006

- Look at from p. 155–166: 19, 34, 38
- Look at from p. 195–201: 19, 24
- Hand in from p. 155–166: 14, 18 (Hint: see picture given in class.), 22a-b
- Hand in from p. 195–201: 1a,b,d, 12, 13, 33 and 18 (Also assume $\mathbb{E}X_n = 0$)*
- Hand in lecture note exercises: 9.1.

* For Problem 18, please add the missing assumption that the random variables should have mean zero. (The assertion to prove is false without this assumption.) With this assumption, $\mathrm{Var}(X) = \mathbb{E}[X^2]$. Also note that $\mathrm{Cov}(X, Y) = 0$ is equivalent to $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$.

0.9 Homework 9. Due Noon, on Wednesday, December 6, 2006

- Look at from p. 195–201: 3, 4, 14, 16, 17, 27, 30
- Hand in from p. 195–201: 15 (Hint: $|a-b| = 2(a-b)^+ - (a-b)$.)
- Hand in from p. 234–246: 1, 2 (Hint: it is just as easy to prove a.s. convergence), 15

Part I

Background Material

1

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* $c \in \mathbb{R}$, $\lim_{n\to\infty} (ca_n) = c \lim_{n\to\infty} a_n$.
3. *If* $\{a_n + b_n\}_{n=1}^{\infty}$ *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

[1] Here we use "a.a. n" as an abreviation 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 $n \geq N$.

$a_n = n$ 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 n is arbitrary, it follows that $\lim_{n\to\infty} (a_n + b_n) = a + b$.

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

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

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

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

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

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

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. $\liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n$ *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* $\lim_{k\to\infty} a_{n_k} = \limsup_{n\to\infty} a_n$. *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. $$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \tag{1.5}$$
whenever the right side of this equation is not of the form $\infty - \infty$.
4. *If* $a_n \ge 0$ *and* $b_n \ge 0$ *for all* $n \in \mathbb{N}$, *then*
$$\limsup_{n\to\infty}(a_n b_n) \le \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n, \tag{1.6}$$
provided the right hand side of (1.6) is not of the form $0 \cdot \infty$ *or* $\infty \cdot 0$.

Proof. Item 1. will be proved here leaving the remaining items as an exercise to the reader. Since

$$\inf\{a_k : k \ge n\} \le \sup\{a_k : k \ge n\} \ \forall\, n,$$

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

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

$$a - \varepsilon \le \inf\{a_k : k \ge N\} \le \sup\{a_k : k \ge N\} \le a + \varepsilon,$$

i.e.

$$a - \varepsilon \le a_k \le a + \varepsilon \text{ for all } k \ge N.$$

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

$$M \le \inf\{a_k : k \ge N\}$$

and hence $\lim_{n\to\infty} a_n = \infty$. The case where $\limsup_{n\to\infty} a_n = -\infty$ 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 $|A - a_n| \le \varepsilon$ for all $n \ge N(\varepsilon)$, i.e.

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

From this we learn that

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

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

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

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

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

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

Proposition 1.6 (Tonelli's theorem for sums). *If $\{a_{kn}\}_{k,n=1}^{\infty}$ is **any** sequence of non-negative numbers, 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. 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 $M \leq L$. Thus we have shown

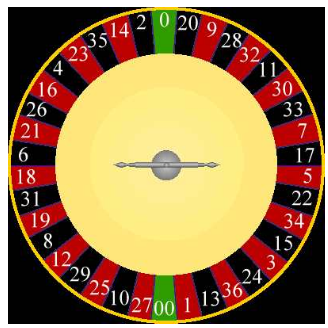
$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = M.$$

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. ∎

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 40 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, $0 \leq t < \infty$. 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 is a subset of Ω.*

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. $A = \cap_{N=1}^{\infty} \cup_{n \geq N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where there are infinitely many heads tossed in the sequence.
4. $A = \cup_{N=1}^{\infty} \cap_{n \geq N} \{\omega \in \Omega : \omega_n = 1\}$ 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 $n \geq N$.

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
$$P(A) = \lim_{N\to\infty} \frac{1}{N} \# \{1 \leq k \leq N : \omega(k) \in A\}. \tag{2.1}$$
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 now that B has occurred. Thus we would like to compute:
$$P(A|B) = \lim_{n\to\infty} \frac{\#\{k : 1 \leq k \leq n \text{ and } \omega_k \in A \cap B\}}{\#\{k : 1 \leq k \leq n \text{ and } \omega_k \in B\}},$$
which represents the frequency that A occurs given that we know that B has occurred. This may be rewritten as
$$\begin{aligned} P(A|B) &= \lim_{n\to\infty} \frac{\frac{1}{n}\#\{k : 1 \leq k \leq n \text{ and } \omega_k \in A \cap B\}}{\frac{1}{n}\#\{k : 1 \leq k \leq n \text{ and } \omega_k \in B\}} \\ &= \frac{P(A \cap B)}{P(B)}. \end{aligned}$$

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. $P(\emptyset) = 1$ and $P(\Omega) = 1$.
3. **Additivity.** If A and B are disjoint event, i.e. $A \cap B = AB = \emptyset$, then

$$\begin{aligned} P(A \cup B) &= \lim_{N \to \infty} \frac{1}{N} \# \{1 \le k \le N : \omega(k) \in A \cup B\} \\ &= \lim_{N \to \infty} \frac{1}{N} [\# \{1 \le k \le N : \omega(k) \in A\} + \# \{1 \le k \le N : \omega(k) \in B\}] \\ &= P(A) + P(B). \end{aligned}$$

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

$$P(\{\omega \in \Omega : (\omega_1, \dots, \omega_k) = \sigma\}) = \frac{1}{2^k} \tag{2.2}$$

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. (2.2) holds, we would like to compute, for example, the probability of the event B where an infinite number of heads are tossed. To try to compute this, let

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

and

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

Since

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

we see that

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

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

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

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

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

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

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

$$zN := \{zn \in S : n \in N\}, \tag{2.3}$$

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

$$P(zN) = P(N) \tag{2.4}$$

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

$$S = \sum_{r \in R} (rN) \tag{2.5}$$

where $\sum_\alpha A_\alpha$ is used to denote the union of pair-wise disjoint sets $\{A_\alpha\}$. By Eqs. (2.4) and (2.5),

$$1 = P(S) = \sum_{r \in R} P(rN) = \sum_{r \in R} P(N). \tag{2.6}$$

We have thus arrived at a contradiction, since the right side of Eq. (2.6) is either equal to 0 or to ∞ depending on whether $P(N) = 0$ or $P(N) > 0$. ■

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

Part II

Formal Development

3

Preliminaries

3.1 Set Operations

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

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

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

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

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

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

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

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

For $A \in 2^X$ let

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

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

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

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

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

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

$$\begin{aligned} \inf_{k \geq n} A_n &:= \cap_{k \geq n} A_k, \\ \sup_{k \geq n} A_n &:= \cup_{k \geq n} A_k, \\ \limsup_{n \to \infty} A_n &:= \{A_n \text{ i.o.}\} := \{x \in X : \#\{n : x \in A_n\} = \infty\} \\ &\text{and} \\ \liminf_{n \to \infty} A_n &:= \{A_n \text{ a.a.}\} := \{x \in X : x \in A_n \text{ for all } n \text{ sufficiently large}\}. \end{aligned}$$

(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 characteristic function of A.

Lemma 3.6. *We have:*

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

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. ■

We end this section with some notation which will be used frequently in the sequel.

Notation 3.9 *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 3.10. *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*

$$\mathcal{E}_A = i_A^{-1}(\mathcal{E}) = \{A \cap E : E \in \mathcal{E}\}.$$

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 3.11. 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 3.12. *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 3.13. *A collection of subsets $\mathcal{A}$ of a set X is an **algebra (Field)** if*

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

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

Definition 3.14. *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 3.15. 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 3.16. *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 3.17. Suppose $X = \{1, 2, 3\}$ and $\mathcal{E} = \{\emptyset, X, \{1, 2\}, \{1, 3\}\}$, see Figure 3.1. Then

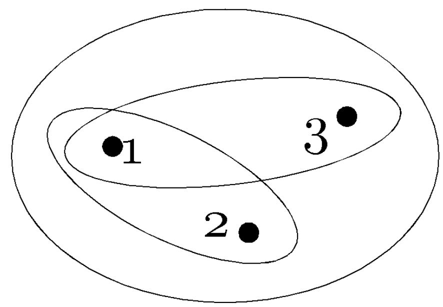


Fig. 3.1. A collection of subsets.

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

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 $\mathcal{E}_1 \neq \mathcal{E}_2$.

Definition 3.18. *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 3.19. Let X be a set and $\mathcal{E} = \{A_1, \ldots, A_n\}$ where $A_1, \ldots, A_n$ is a partition of X. In this case

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

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

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

Example 3.20. Suppose that X is a finite set and that $\mathcal{A} \subset 2^X$ is an algebra. 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. Let $C = A_x \cap A_y \in \mathcal{A}$. I claim that if $C \neq \emptyset$, then $A_x = A_y$. To see this, let us first consider the case where $\{x, y\} \subset C$. In this case we must have $A_x \subset C$ and $A_y \subset C$ and therefore $A_x = A_y$. Now suppose either x or y is not in C. For definiteness, say $x \notin C$, i.e. $x \notin y$. Then $x \in A_x \setminus A_y \in \mathcal{A}$ from which it follows that $A_x = A_x \setminus A_y$, i.e. $A_x \cap A_y = \emptyset$.

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

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

Proposition 3.21. *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 3.20. 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 3.22 (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 3.7. 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 $\Lambda = \emptyset$.

Proposition 3.23. *Let X be a set and $\mathcal{E} \subset 2^X$. Let $\mathcal{E}^c := \{A^c : A \in \mathcal{E}\}$ and $\mathcal{E}_c := \mathcal{E} \cup \{X, \emptyset\} \cup \mathcal{E}^c$ Then*

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

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

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

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

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

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 3.24. 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 3.21.

Exercise 3.8. 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 (3.8). 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 3.23.

Definition 3.25. *The Borel σ – field, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$, on $\mathbb{R}$ is the smallest σ -field containing all of the open subsets of* $\mathbb{R}$.

Exercise 3.9. Verify the 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.

Exercise 3.10. Suppose $f : X \to Y$ is a function, $\mathcal{F} \subset 2^Y$ and $\mathcal{B} \subset 2^X$. Show $f^{-1}\mathcal{F}$ and $f_*\mathcal{B}$ (see Notation 3.9) are algebras (σ – algebras) provided $\mathcal{F}$ and $\mathcal{B}$ are algebras (σ – algebras).

Lemma 3.26. *Suppose that $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ and $A \subset Y$ then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E}))\ \text{and} \tag{3.3}$$

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A\ ), \tag{3.4}$$

where $\mathcal{B}_A := \{B \cap A : B \in \mathcal{B}\}$. (Similar assertion hold with $\sigma(\cdot)$ being replaced by $\mathcal{A}(\cdot)$.)

Proof. By Exercise 3.10, $f^{-1}(\sigma(\mathcal{E}))$ is a σ – algebra and since $\mathcal{E} \subset \mathcal{F}$, $f^{-1}(\mathcal{E}) \subset f^{-1}(\sigma(\mathcal{E}))$. It now follows that

$$\sigma(f^{-1}(\mathcal{E})) \subset f^{-1}(\sigma(\mathcal{E})).$$

For the reverse inclusion, notice that

$$f_*\sigma\left(f^{-1}(\mathcal{E})\right) := \left\{B \subset Y : f^{-1}(B) \in \sigma\left(f^{-1}(\mathcal{E})\right)\right\}$$

is a σ – algebra which contains $\mathcal{E}$ and thus $\sigma(\mathcal{E}) \subset f_*\sigma\left(f^{-1}(\mathcal{E})\right)$. Hence for every $B \in \sigma(\mathcal{E})$ we know that $f^{-1}(B) \in \sigma\left(f^{-1}(\mathcal{E})\right)$, i.e.

$$f^{-1}(\sigma(\mathcal{E})) \subset \sigma\left(f^{-1}(\mathcal{E})\right).$$

Applying Eq. (3.3) with $X = A$ and $f = i_A$ being the inclusion map implies

$$(\sigma(\mathcal{E}))_A = i_A^{-1}(\sigma(\mathcal{E})) = \sigma(i_A^{-1}(\mathcal{E})) = \sigma(\mathcal{E}_A).$$

■

Example 3.27. Let $\mathcal{E} = \{(a,b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel $\sigma-$ 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 $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

Definition 3.28. *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 3.29 *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 3.30. *For every algebra $\mathcal{A} \subset 2^{\Omega}$, the set simple random variables, $\mathbb{S}(\mathcal{A})$, forms an algebra.*

Proof. 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

$$\{f + cg = \lambda\} = \bigcup_{a,b \in \mathbb{C} : a + cb = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{3.5}$$

and

$$\{f \cdot g = \lambda\} = \bigcup_{a,b \in \mathbb{C} : a \cdot b = \lambda} (\{f = a\} \cap \{g = b\}) \in \mathcal{A} \tag{3.6}$$

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

Definition 3.31. *A **simple function algebra**, $\mathbb{S}$, is a subalgebra of the bounded complex functions on X such that $1 \in \mathbb{S}$ and each function, $f \in \mathbb{S}$, 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.)

Lemma 3.32. *Suppose that $\mathbb{S}$ is a simple function algebra, $f \in \mathbb{S}$ and $\alpha \in f(X)$. 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 $\{f = \alpha\} \in \mathcal{A}$. ■

Exercise 3.11. 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 (3.11).

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 3.32, 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

Finitely Additive Measures

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* ***monotonic*** *if* $\mu(A) \le \mu(B)$ *for all* $A, B \in \mathcal{E}$ *with* $A \subset B$.
2. 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* $n \in \mathbb{N} \cup \{\infty\}$ $(n \in \mathbb{N})$.
3. μ *is* ***super-additive (finitely super-additive) on*** $\mathcal{E}$ *if*
$$\mu(E) \ge \sum_{i=1}^{n} \mu(E_i) \tag{4.1}$$
whenever $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ *with* $n \in \mathbb{N} \cup \{\infty\}$ $(n \in \mathbb{N})$.
4. μ *is* ***additive or finitely additive on*** $\mathcal{E}$ *if*
$$\mu(E) = \sum_{i=1}^{n} \mu(E_i) \tag{4.2}$$
whenever $E = \sum_{i=1}^{n} E_i \in \mathcal{E}$ *with* $E_i \in \mathcal{E}$ *for* $i = 1, 2, \dots, n < \infty$.
5. *If* $\mathcal{E} = \mathcal{A}$ *is an algebra,* $\mu(\emptyset) = 0$, *and* μ *is finitely additive on* $\mathcal{A}$, *then* μ *is said to be a* ***finitely additive measure.***
6. μ *is* σ – ***additive (or countable additive) on*** $\mathcal{E}$ *if item 4. holds even when* $n = \infty$.
7. *If* $\mathcal{E} = \mathcal{A}$ *is an algebra,* $\mu(\emptyset) = 0$, *and* μ *is* σ – *additive on* $\mathcal{A}$ *then* μ *is called a* ***premeasure on*** $\mathcal{A}$.
8. *A* ***measure*** *is a premeasure,* $\mu : \mathcal{B} \to [0, \infty]$, *where* $\mathcal{B}$ *is a* σ – *algebra. We say that* μ *is a* ***probability measure*** *if* $\mu(X) = 1$.

4.1 Finitely Additive Measures

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

1. *(μ is **monotone**) $\mu(E) \leq \mu(F)$ if $E \subset F$.*
2. *For $A, B \in \mathcal{A}$, the following **strong additivity formula** holds;*

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{4.3}$$

3. *(μ is **finitely subbadditive**) $\mu(\cup_{j=1}^n E_j) \leq \sum_{j=1}^n \mu(E_j)$.*
4. *μ is sub-additive on $\mathcal{A}$ iff*

$$\mu(A) \leq \sum_{i=1}^\infty \mu(A_i) \text{ for } A = \sum_{i=1}^\infty A_i \tag{4.4}$$

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*

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

6. *A finitely additive measure, μ, is a premeasure iff μ is sub-additve.*

Proof.

1. Since F is the disjoint union of E and $(F \setminus E)$ and $F \setminus E = F \cap E^c \in \mathcal{A}$ it follows that
$$\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E).$$
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. (4.3).
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
$$\mu(E) = \sum \mu(\widetilde{E}_j) \leq \sum \mu(E_j).$$
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. (4.4)
$$\mu(A) \leq \sum_{i=1}^\infty \mu(A_i) \leq \sum_{i=1}^\infty \mu(B_i).$$
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, $\sum_{i=1}^n \mu(A_i) \leq \mu(A)$. Letting $n \to \infty$ in this equation shows μ is superadditive.

6. This is a combination of items 5. and 6.

■

Proposition 4.3. *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 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 $A_n \downarrow \Omega$, $P(A_n) \downarrow 1$.*

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,

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

It is clear that $2 \implies 4$ and that $3 \implies 5$. To finish the proof we will show $5 \implies 2$ and $5 \implies 3$.

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

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

$5 \implies 3$. If $A_n \in \mathcal{A}$ such that $A_n \downarrow A \in \mathcal{A}$, then $A_n \setminus A \downarrow \emptyset$. Therefore,

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

■

Remark 4.4. 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.

Definition 4.5. *Let $(\Omega, \mathcal{B})$ be a* ***measurable space****, i.e. $\mathcal{B} \subset 2^{\Omega}$ is a σ – algebra. A* ***probability measure on*** *$(\Omega, \mathcal{B})$ is a finitely additive probability measure, $P : \mathcal{B} \to [0,1]$ such that any and hence all of the continuity properties in Proposition 4.3 hold. We will call $(\Omega, \mathcal{B}, P)$ a probability space.*

Lemma 4.6. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, then P is countably sub-additive.*

Proof. Suppose that $A_n \in \mathcal{B}$ and let $A_1' := A_1$ and for $n \geq 2$, let $A_n' := A_n \setminus (A_1 \cup \dots A_{n-1}) \in \mathcal{B}$. Then

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

■

4.2 Examples of Measures

Most σ – 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 4.7. Suppose that Ω is a finite set, $\mathcal{B} := 2^{\Omega}$, and $p : \Omega \to [0,1]$ is a function such that

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

Then

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

defines a measure on 2^{Ω}.

Example 4.8. 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 4.9. Suppose that μ is a measure on X and $\lambda > 0$, then $\lambda \cdot \mu$ is also a measure on X. Moreover, if $\{\mu_j\}_{j \in J}$ are all measures on X, then $\mu = \sum_{j=1}^{\infty} \mu_j$, i.e.

$$\mu(A) = \sum_{j=1}^{\infty} \mu_j(A) \text{ for all } A \subset X$$

is a measure on X. (See Section 3.1 for the meaning of this sum.) To prove this we must show that μ is countably additive. Suppose that $\{A_i\}_{i=1}^{\infty}$ is a collection of pair-wise disjoint subsets of X, then

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

wherein the third equality we used Theorem 1.6 and in the fourth we used that fact that μ_j is a measure.

Example 4.10. Suppose that X is a set $\lambda : X \to [0, \infty]$ is a function. Then

$$\mu := \sum_{x \in X} \lambda(x) \delta_x$$

is a measure, explicitly

$$\mu(A) = \sum_{x \in A} \lambda(x)$$

for all $A \subset X$.

Example 4.11. Suppose that $\mathcal{F} \subset 2^X$ is a countable or finite partition of X and $\mathcal{B} \subset 2^X$ is the $\sigma -$ algebra which consists of the collection of sets $A \subset X$ such that

$$A = \cup \{\alpha \in \mathcal{F} : \alpha \subset A\}. \tag{4.5}$$

Any measure $\mu : \mathcal{B} \to [0, \infty]$ is determined uniquely by its values on $\mathcal{F}$. Conversely, if we are given any function $\lambda : \mathcal{F} \to [0, \infty]$ we may define, for $A \in \mathcal{B}$,

$$\mu(A) = \sum_{\alpha \in \mathcal{F} \ni \alpha \subset A} \lambda(\alpha) = \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A}$$

where $1_{\alpha \subset A}$ is one if $\alpha \subset A$ and zero otherwise. We may check that μ is a measure on $\mathcal{B}$. Indeed, if $A = \sum_{i=1}^{\infty} A_i$ and $\alpha \in \mathcal{F}$, then $\alpha \subset A$ iff $\alpha \subset A_i$ for one and hence exactly one A_i. Therefore $1_{\alpha \subset A} = \sum_{i=1}^{\infty} 1_{\alpha \subset A_i}$ and hence

$$\begin{aligned}\mu(A) &= \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A} = \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) \sum_{i=1}^{\infty} 1_{\alpha \subset A_i} \\ &= \sum_{i=1}^{\infty} \sum_{\alpha \in \mathcal{F}} \lambda(\alpha) 1_{\alpha \subset A_i} = \sum_{i=1}^{\infty} \mu(A_i)\end{aligned}$$

as desired. Thus we have shown that there is a one to one correspondence between measures μ on $\mathcal{B}$ and functions $\lambda : \mathcal{F} \to [0, \infty]$.

The following example explains what is going on in a more typical case of interest to us in the sequel.

Example 4.12. Suppose that $\Omega = \mathbb{R}$, $\mathcal{A}$ consists of those sets, $A \subset \mathbb{R}$ which may be written as finite disjoint unions from

$$\mathcal{S} := \{(a, b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}.$$

We will show below the following:

1. $\mathcal{A}$ is an algebra. (Recall that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A})$.)
2. To every increasing function, $F : \mathbb{R} \to [0, 1]$ such that

$$F(-\infty) := \lim_{x \to -\infty} F(x) = 0 \text{ and}$$
$$F(+\infty) := \lim_{x \to \infty} F(x) = 1$$

there exists a finitely additive probability measure, $P = P_F$ on $\mathcal{A}$ such that

$$P((a, b] \cap \mathbb{R}) = F(b) - F(a) \text{ for all } -\infty \leq a \leq b \leq \infty.$$

3. P is σ – additive on $\mathcal{A}$ iff F is right continuous.
4. P extends to a probability measure on $\mathcal{B}_{\mathbb{R}}$ iff F is right continuous.

Let us observe directly that if $F(a+) := \lim_{x \downarrow a} F(x) \neq F(a)$, then $(a, a+1/n] \downarrow \emptyset$ while

$$P((a, a + 1/n]) = F(a + 1/n) - F(a) \downarrow F(a+) - F(a) > 0.$$

Hence P can not be σ – additive on $\mathcal{A}$ in this case.

4.3 Simple Integration

Definition 4.13 (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*

$$\mathbb{E}_P(f) = \sum_{y \in \mathbb{C}} y P(f = y). \tag{4.6}$$

Example 4.14. Suppose that $A \in \mathcal{A}$, then

$$\mathbb{E}1_A = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A). \tag{4.7}$$

Remark 4.15. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (2.1) by

$$P(A) = \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \omega(k) \in A\}$$

where $\omega(k) \in \Omega$ was the result of the k^{th} "independent" experiment. If we use this interpretation back in Eq. (4.6), we arrive at

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

Thus informally, $\mathbb{E}f$ should represent the average of the values of f over many "independent" experiments.

Proposition 4.16. *The expectation operator,* $\mathbb{E} = \mathbb{E}_P$, *satisfies:*

1. *If* $f \in \mathbb{S}(\mathcal{A})$ *and* $\lambda \in \mathbb{C}$, *then*
$$\mathbb{E}(\lambda f) = \lambda \mathbb{E}(f). \tag{4.8}$$
2. *If* $f, g \in \mathbb{S}(\mathcal{A})$, *then*
$$\mathbb{E}(f+g) = \mathbb{E}(g) + \mathbb{E}(f). \tag{4.9}$$
3. $\mathbb{E}$ *is* ***positive****, i.e.* $\mathbb{E}(f) \ge 0$ *if* f *is a non-negative measurable simple function.*
4. *For all* $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}f| \le \mathbb{E}|f|. \tag{4.10}$$

Proof.

1. If $\lambda \ne 0$, then

$$\begin{aligned}
\mathbb{E}(\lambda f) &= \sum_{y\in\mathbb{C}\cup\{\infty\}} y\ P(\lambda f = y) = \sum_{y\in\mathbb{C}\cup\{\infty\}} y\ P(f = y/\lambda) \\
&= \sum_{z\in\mathbb{C}\cup\{\infty\}} \lambda z\ P(f=z) = \lambda \mathbb{E}(f).
\end{aligned}$$

The case $\lambda = 0$ is trivial.

2. Writing $\{f = a, g = b\}$ for $f^{-1}(\{a\}) \cap g^{-1}(\{b\})$, then

$$\begin{aligned}
\mathbb{E}(f+g) &= \sum_{z\in\mathbb{C}} z\ P(f+g=z) \\
&= \sum_{z\in\mathbb{C}} z\ P\left(\cup_{a+b=z}\{f=a,\ g=b\}\right) \\
&= \sum_{z\in\mathbb{C}} z \sum_{a+b=z} P\left(\{f=a,\ g=b\}\right) \\
&= \sum_{z\in\mathbb{C}} \sum_{a+b=z} (a+b)\, P\left(\{f=a,\ g=b\}\right) \\
&= \sum_{a,b} (a+b)\, P\left(\{f=a,\ g=b\}\right).
\end{aligned}$$

But

$$\begin{aligned}
\sum_{a,b} aP\left(\{f=a,\ g=b\}\right) &= \sum_a a \sum_b P\left(\{f=a,\ g=b\}\right) \\
&= \sum_a aP\left(\cup_b \{f=a,\ g=b\}\right) \\
&= \sum_a aP\left(\{f=a\}\right) = \mathbb{E}f
\end{aligned}$$

and similarly,

$$\sum_{a,b} bP\left(\{f=a,\ g=b\}\right) = \mathbb{E}g.$$

Equation (4.9) is now a consequence of the last three displayed equations.

3. If $f \geq 0$ then

$$\mathbb{E}(f) = \sum_{a\geq 0} aP(f=a) \geq 0.$$

4. First observe that

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

and therefore,

$$\mathbb{E}\,|f| = \mathbb{E}\sum_{\lambda\in\mathbb{C}} |\lambda|\, 1_{f=\lambda} = \sum_{\lambda\in\mathbb{C}} |\lambda|\, \mathbb{E}1_{f=\lambda} = \sum_{\lambda\in\mathbb{C}} |\lambda|\, P\left(f=\lambda\right) \leq \max|f|\,.$$

On the other hand,

$$|\mathbb{E}f| = \left|\sum_{\lambda\in\mathbb{C}} \lambda P\left(f=\lambda\right)\right| \leq \sum_{\lambda\in\mathbb{C}} |\lambda|\, P\left(f=\lambda\right) = \mathbb{E}\,|f|\,.$$

■

Remark 4.17. Every simple measurable function, $f : \Omega \to \mathbb{C}$, may be written as $f = \sum_{j=1}^{N} \lambda_j 1_{A_j}$ for some $\lambda_j \in \mathbb{C}$ and some $A_j \in \mathbb{C}$. Moreover if f is represented this way, then

$$\mathbb{E}f = \mathbb{E}\left[\sum_{j=1}^{N} \lambda_j 1_{A_j}\right] = \sum_{j=1}^{N} \lambda_j \mathbb{E} 1_{A_j} = \sum_{j=1}^{N} \lambda_j P(A_j).$$

Remark 4.18 (Chebyshev's Inequality). Suppose that $f \in \mathbb{S}(\mathcal{A})$, $\varepsilon > 0$, and $p > 0$, then

$$P(\{|f| \geq \varepsilon\}) = \mathbb{E}\left[1_{|f|\geq\varepsilon}\right] \leq \mathbb{E}\left[\frac{|f|^p}{\varepsilon^p} 1_{|f|\geq\varepsilon}\right] \leq \varepsilon^{-p}\mathbb{E}|f|^p. \tag{4.11}$$

Observe that

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

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

Lemma 4.19 (Inclusion Exclusion Formula). *If $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots, M$ such that $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, then*

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^{M} (-1)^{k+1} \sum_{1\leq n_1<n_2<\cdots<n_k\leq M} \mu(A_{n_1} \cap \cdots \cap A_{n_k}). \tag{4.12}$$

Proof. This may be proved inductively from Eq. (4.3). 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

$$\begin{aligned} 1 - 1_A = 1_{A^c} &= \prod_{n=1}^{M} 1_{A_n^c} = \prod_{n=1}^{M} (1 - 1_{A_n}) \\ &= \sum_{k=0}^{M} (-1)^k \sum_{0\leq n_1<n_2<\cdots<n_k\leq M} 1_{A_{n_1}} \cdots 1_{A_{n_k}} \\ &= \sum_{k=0}^{M} (-1)^k \sum_{0\leq n_1<n_2<\cdots<n_k\leq M} 1_{A_{n_1}\cap\cdots\cap A_{n_k}} \end{aligned}$$

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^{M} (-1)^{k+1} \sum_{1\leq n_1<n_2<\cdots<n_k\leq M} 1_{A_{n_1}\cap\cdots\cap A_{n_k}}. \tag{4.13}$$

Taking expectations of this equation then gives Eq. (4.12). ■

Remark 4.20. Here is an alternate proof of Eq. (4.13). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1 \cap \cdots \cap A_m$ and $\omega \notin A_{m+1} \cup \cdots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (4.13) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

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

This verifies Eq. (4.13) since $1_{\cup_{n=1}^M A_n} (\omega) = 1$.

Example 4.21 (Coincidences). Let Ω be the set of permutations (think of card shuffling), $\omega : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$, 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

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

Since

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

and

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

we find

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

For large n this gives,

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

Example 4.22. Continue the notation in Example 4.21. 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 formula when $n = 6$. 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

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

while

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

and

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

4.4 Simple Independence and the Weak Law of Large Numbers

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

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

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

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \in 2^{\Omega}. \tag{4.14}$$

Exercise 4.1 (Simple Independence 1.). Suppose $q_i : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda\in\Lambda} q_i(\lambda) = 1$ for $i = 1, 2, \ldots, n$ and If $p(\omega) = \prod_{i=1}^{n} q_i(\omega_i)$. 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 $Q_i(\gamma) = \sum_{\lambda\in\gamma} q_i(\lambda)$ for all $\gamma \subset \Lambda$.

Exercise 4.2 (Simple Independence 2.). Prove the converse of the previous exercise. Namely, if

$$\mathbb{E}_P\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}_P[f_i(X_i)] \tag{4.15}$$

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 $p(\omega) = \prod_{i=1}^{n} q_i(\omega_i)$.

Exercise 4.3 (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. (4.14). Further let $X_i(\omega) = \omega_i$ for $i = 1, 2, \ldots, n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

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

1. Show, $\xi = \sum_{\lambda\in\Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda\in\Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda\in\Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{4.16}$$
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
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{4.17}$$

5. Conclude using Eq. (4.17) and Remark 4.18 that
$$P\left(|S_n-\xi|\geq\varepsilon\right)\leq\frac{1}{n\varepsilon^2}\sigma^2. \tag{4.18}$$
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.

Exercise 4.4 (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,
$$\xi(x):=\mathbb{E}_QX=x \text{ and}$$
$$\sigma^2(x):=\mathbb{E}_Q(X-x)^2=(1-x)x\leq 1/4.$$

Theorem 4.23 (Weierstrass Approximation Theorem via Bernstein's Polynomials.). *Suppose that* $f\in C([0,1],\mathbb{C})$ *and*
$$p_n(x):=\sum_{k=0}^n\binom{n}{k}f\left(\frac{k}{n}\right)x^k(1-x)^{n-k}.$$
Then
$$\lim_{n\to\infty}\sup_{x\in[0,1]}|f(x)-p_n(x)|=0.$$
(See Theorem 14.42 for a multi-dimensional generalization of this theorem.)

Proof. Let $x\in[0,1]$, $\Lambda=\{0,1\}$, $q(0)=1-x$, $q(1)=x$, $\Omega=\Lambda^n$, and
$$P_x(\{\omega\})=q(\omega_1)\dots q(\omega_n)=x^{\sum_{i=1}^n\omega_i}\cdot(1-x)^{1-\sum_{i=1}^n\omega_i}.$$
As above, let $S_n=\frac{1}{n}(X_1+\dots+X_n)$, where $X_i(\omega)=\omega_i$ and observe that
$$P_x\left(S_n=\frac{k}{n}\right)=\binom{n}{k}x^k(1-x)^{n-k}.$$
Therefore, writing $\mathbb{E}_x$ for $\mathbb{E}_{P_x}$, we have
$$\mathbb{E}_x[f(S_n)]=\sum_{k=0}^nf\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}=p_n(x).$$
Hence we find
$$\begin{aligned}|p_n(x)-f(x)|&=|\mathbb{E}_xf(S_n)-f(x)|=|\mathbb{E}_x[f(S_n)-f(x)]|\\&\leq\mathbb{E}_x|f(S_n)-f(x)|\\&=\mathbb{E}_x[|f(S_n)-f(x)|:|S_n-x|\geq\varepsilon]\\&\qquad+\mathbb{E}_x[|f(S_n)-f(x)|:|S_n-x|<\varepsilon]\\&\leq 2M\cdot P_x(|S_n-x|\geq\varepsilon)+\delta(\varepsilon)\end{aligned}$$

where

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

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

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

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 $\varepsilon \downarrow 0$.

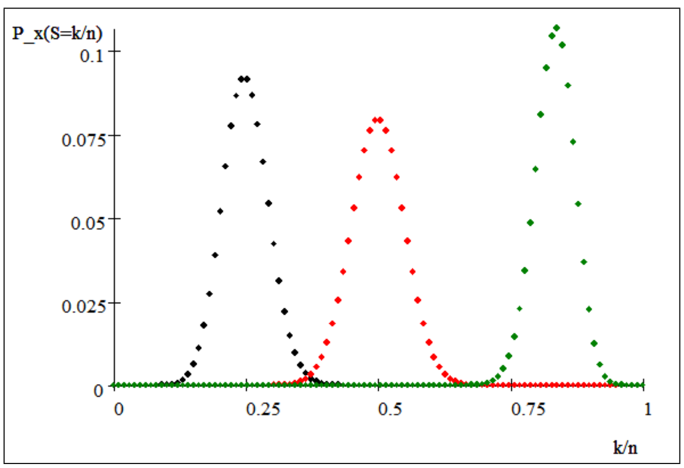


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

■

4.5 Constructing Finitely Additive Measures

Definition 4.24. *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 $\mathcal{S}$.)*

Example 4.25. Let $X = \mathbb{R}$, then

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

is a semi-field

Exercise 4.5. Let $\mathcal{A} \subset 2^X$ and $\mathcal{B} \subset 2^Y$ be semi-fields. Show the collection

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

is also a semi-field.

Proposition 4.26. *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. 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 $\emptyset, X \in \mathcal{A}$. Now suppose that $A_i = \sum_{F \in \Lambda_i} F \in \mathcal{A}$ where, for $i = 1, 2, \ldots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{S}$. Then

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

and this is a disjoint (you check) union of elements from $\mathcal{S}$. Therefore $\mathcal{A}$ is closed under finite intersections. Similarly, 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 $F \in \Lambda \subset \mathcal{S}$ and $\mathcal{A}$ is closed under finite intersections, it follows that $A^c \in \mathcal{A}$. ■

Example 4.27. Let $X = \mathbb{R}$ and $\mathcal{S} := \left\{(a,b] \cap \mathbb{R} : a,b \in \bar{\mathbb{R}}\right\}$ be as in Example 4.25. Then $\mathcal{A}(\mathcal{S})$ may be described as being those sets which are finite disjoint unions of sets from $\mathcal{S}$.

Proposition 4.28 (Construction of Finitely Additive Measures). *Suppose $\mathcal{S} \subset 2^X$ is a semi-algebra (see Definition 4.24) and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ is the algebra generated by $\mathcal{S}$. Then every additive function $\mu : \mathcal{S} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ extends uniquely to an additive measure (which we still denote by μ) on $\mathcal{A}$.*

Proof. Since (by Proposition 4.26) 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

$$\mu(A) = \sum_i \mu(E_i). \tag{4.19}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (4.19) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

$$\sum_i \mu(E_i) = \sum_j \mu(F_j). \tag{4.20}$$

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. (4.20) holds. It is now easy to verify that μ extended to $\mathcal{A}$ as in Eq. (4.19) is an additive measure on $\mathcal{A}$. ■

Proposition 4.29. *Let $X = \mathbb{R}$, $\mathcal{S}$ be a semi-algebra*

$$\mathcal{S} = \{(a, b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{4.21}$$

and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ be the algebra formed by taking finite disjoint unions of elements from $\mathcal{S}$, see Proposition 4.26. 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

$$\mu((a, b] \cap \mathbb{R}) = F(b) - F(a) \ \forall\ a \le b \text{ in } \bar{\mathbb{R}}. \tag{4.22}$$

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. (4.22) holds.

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. (4.22). The argument will be completed by showing μ is additive on $\mathcal{S}$ and hence, by Proposition 4.28, 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,

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

■

5

Countably Additive Measures

5.1 Distribution Function for Probability Measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$

Definition 5.1. *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*

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

Example 5.2. 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,

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

Lemma 5.3. *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. $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$,
2. $F(\infty) := \lim_{x\to\infty} F(x) = 1$,
3. *F is non-decreasing, and*
4. *F is right continuous.*

Theorem 5.4. *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 4. in Lemma 5.3, 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 in the theorem is covered in Exercise 5.1 below. The existence portion of the Theorem follows from Proposition 5.7 and Theorem 5.19 below. ■

Example 5.5 (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 $(0,1]$.

Recall from Definition 3.14 that $\mathcal{B} \subset 2^X$ is a σ – algebra on X if $\mathcal{B}$ is an algebra which is closed under countable unions and intersections.

5.2 Construction of Premeasures

Proposition 5.6. *Suppose that $\mathcal{S} \subset 2^X$ 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 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 $A = \sum_{n=1}^{\infty} A_n$ with $A \in \mathcal{A}$ and each $A_n \in \mathcal{A}$ which we express as $A = \sum_{j=1}^{k} E_j$ with $E_j \in \mathcal{S}$ and $A_n = \sum_{i=1}^{N_n} E_{n,i}$ with $E_{n,i} \in \mathcal{S}$. Then

$$E_j = A \cap E_j = \sum_{n=1}^{\infty} A_n \cap E_j = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} E_{n,i} \cap E_j$$

which is a countable union and hence by assumption,

$$\mu(E_j) \leq \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j).$$

Summing this equation on j and using the finite additivity of μ shows

$$\begin{aligned} \mu(A) = \sum_{j=1}^{k} \mu(E_j) &\leq \sum_{j=1}^{k} \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i} \cap E_j) \\ &= \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \sum_{j=1}^{k} \mu(E_{n,i} \cap E_j) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu(E_{n,i}) = \sum_{n=1}^{\infty} \mu(A_n), \end{aligned}$$

which proves (using Proposition 4.2) the sub-additivity of μ on $\mathcal{A}$. ■

Now suppose that $F : \mathbb{R} \to \mathbb{R}$ be an increasing function, $F(\pm\infty) := \lim_{x \to \pm\infty} F(x)$ and $\mu = \mu_F$ be the finitely additive measure on $(\mathbb{R}, \mathcal{A})$ described in Proposition 4.29. If μ happens to be a premeasure on $\mathcal{A}$, then, letting $A_n = (a, b_n]$ with $b_n \downarrow b$ as $n \to \infty$, implies

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

Since $\{b_n\}_{n=1}^{\infty}$ was an arbitrary sequence such that $b_n \downarrow b$, we have shown $\lim_{y \downarrow b} F(y) = F(b)$, i.e. F is right continuous. The next proposition shows the converse is true as well. Hence premeasures on $\mathcal{A}$ which are finite on bounded sets are in one to one correspondences with right continuous increasing functions which vanish at 0.

Proposition 5.7. *To each right continuous increasing function* $F : \mathbb{R} \to \mathbb{R}$ *there exists a unique premeasure* $\mu = \mu_F$ *on* $\mathcal{A}$ *such that*

$$\mu_F((a, b]) = F(b) - F(a) \ \forall \ -\infty < a < b < \infty.$$

Proof. As above, let $F(\pm\infty) := \lim_{x \to \pm\infty} F(x)$ and $\mu = \mu_F$ be as in Proposition 4.29. Because of Proposition 5.6, to finish the proof it suffices to show μ is sub-additive on $\mathcal{S}$.

First suppose that $-\infty < a < b < \infty$, $J = (a, b]$, $J_n = (a_n, b_n]$ such that $J = \sum_{n=1}^{\infty} J_n$. We wish to show

$$\mu(J) \leq \sum_{n=1}^{\infty} \mu(J_n). \tag{5.1}$$

To do this choose numbers $\tilde{a} > a$, $\tilde{b}_n > b_n$ in which case $I := (\tilde{a}, b] \subset J$,

$$\tilde{J}_n := (a_n, \tilde{b}_n] \supset \tilde{J}_n^o := (a_n, \tilde{b}_n) \supset J_n.$$

Since $\bar{I} = [\tilde{a}, b]$ is compact and $\bar{I} \subset J \subset \bigcup_{n=1}^{\infty} \tilde{J}_n^o$ there exists[1] $N < \infty$ such that

$$I \subset \bar{I} \subset \bigcup_{n=1}^{N} \tilde{J}_n^o \subset \bigcup_{n=1}^{N} \tilde{J}_n.$$

Hence by **finite** sub-additivity of μ,

$$F(b) - F(\tilde{a}) = \mu(I) \leq \sum_{n=1}^{N} \mu(\tilde{J}_n) \leq \sum_{n=1}^{\infty} \mu(\tilde{J}_n).$$

Using the right continuity of F and letting $\tilde{a} \downarrow a$ in the above inequality,

[1] To see this, let $c := \sup\left\{x \leq b : [\tilde{a}, x] \text{ is finitely covered by } \left\{\tilde{J}_n^o\right\}_{n=1}^{\infty}\right\}$. If $c < b$, then $c \in \tilde{J}_m^o$ for some m and there exists $x \in \tilde{J}_m^o$ such that $[\tilde{a}, x]$ is finitely covered by $\left\{\tilde{J}_n^o\right\}_{n=1}^{\infty}$, say by $\left\{\tilde{J}_n^o\right\}_{n=1}^{N}$. We would then have that $\left\{\tilde{J}_n^o\right\}_{n=1}^{\max(m,N)}$ finitely covers $[a, c']$ for all $c' \in \tilde{J}_m^o$. But this contradicts the definition of c.

$$\mu(J) = \mu((a,b]) = F(b) - F(a) \le \sum_{n=1}^{\infty} \mu\left(\tilde{J}_n\right)$$
$$= \sum_{n=1}^{\infty} \mu(J_n) + \sum_{n=1}^{\infty} \mu(\tilde{J}_n \setminus J_n). \tag{5.2}$$

Given $\varepsilon > 0$, we may use the right continuity of F to choose $\tilde{b}_n$ so that

$$\mu(\tilde{J}_n \setminus J_n) = F(\tilde{b}_n) - F(b_n) \le \varepsilon 2^{-n} \ \forall \ n \in \mathbb{N}.$$

Using this in Eq. (5.2) shows

$$\mu(J) = \mu((a,b]) \le \sum_{n=1}^{\infty} \mu(J_n) + \varepsilon$$

which verifies Eq. (5.1) since $\varepsilon > 0$ was arbitrary.

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. ■

Before continuing our development of the existence of measures, we will pause to show that measures are often uniquely determined by their values on a generating sub-algebra. This detour will also have the added benefit of motivating Carathoedory's existence proof to be given below.

5.3 Regularity and Uniqueness Results

Definition 5.8. *Given a collection of subsets, $\mathcal{E}$, of X, let $\mathcal{E}_\sigma$ denote the collection of subsets of X which are finite or countable unions of sets from $\mathcal{E}$. Similarly let $\mathcal{E}_\delta$ denote the collection of subsets of X 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 5.9. *Suppose that $\mathcal{A} \subset 2^X$ 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. ∎

Theorem 5.10 (Finite Regularity Result). *Suppose $\mathcal{A} \subset 2^X$ is an algebra, $\mathcal{B} = \sigma(\mathcal{A})$ and $\mu : \mathcal{B} \to [0, \infty)$ is a finite measure, i.e. $\mu(X) < \infty$. 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 $i = 1, 2, \ldots$ 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,

$$\mu(C \setminus A) = \mu(\cup_{i=1}^\infty [C_i \setminus A]) \leq \sum_{i=1}^\infty \mu(C_i \setminus A) \leq \sum_{i=1}^\infty \mu(C_i \setminus A_i) < \varepsilon.$$

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 5.11. *Suppose X is a set, $\mathcal{E} \subset \mathcal{B} \subset 2^X$ 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 $X = \cup_{n=1}^\infty E_n$. 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 $(X, \mathcal{B}, \mu)$ 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 $X_n \in \mathcal{B}$ such that $X_n \uparrow X$ and $\mu(X_n) < \infty$.

Corollary 5.12 (σ – Finite Regularity Result). *Theorem 5.10 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 $X_n \in \mathcal{A}$ such that $\cup_{n=1}^\infty X_n = X$ and $\mu(X_n) < \infty$ for all n.Since $A \in \mathcal{B} \to \mu_n(A) := \mu(X_n \cap A)$ is a finite measure on $A \in \mathcal{B}$ for each n, by Theorem 5.10, for every $B \in \mathcal{B}$ there exists $C_n \in \mathcal{A}_\sigma$ such that $B \subset C_n$ and $\mu(X_n \cap [C_n \setminus B]) = \mu_n(C_n \setminus B) < 2^{-n}\varepsilon$. Now let $C := \cup_{n=1}^\infty [X_n \cap C_n] \in \mathcal{A}_\sigma$ and observe that $B \subset C$ and

$$\begin{aligned}\mu(C \setminus B) &= \mu\left(\cup_{n=1}^\infty ([X_n \cap C_n] \setminus B)\right)\\ &\leq \sum_{n=1}^\infty \mu([X_n \cap C_n] \setminus B) = \sum_{n=1}^\infty \mu(X_n \cap [C_n \setminus B]) < \varepsilon.\end{aligned}$$

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

$$\mu(C \setminus A) = \mu([B \setminus A] \cup [(C \setminus B) \setminus A]) \leq \mu(B \setminus A) + \mu(C \setminus B) < 2\varepsilon$$

and the result is proved. ■

Exercise 5.1. Suppose $\mathcal{A} \subset 2^X$ 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 5.13. *Suppose $\mathcal{A} \subset 2^X$ 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 5.10, 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 $\mu(C_n \setminus B) \leq 1/n$ and $\mu(B \setminus A_n) \leq 1/n$. 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

$$\begin{aligned}\mu(C \setminus A) &= \mu(C \setminus B) + \mu(B \setminus A) \leq \mu(C_n \setminus B) + \mu(B \setminus A_n) \\ &\leq 2/n \to 0 \text{ as } n \to \infty.\end{aligned}$$

■

Exercise 5.2. Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma\left(\{\text{open subsets of } \mathbb{R}^n\}\right)$ 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 $\mu\left(V \setminus F\right) < \varepsilon$. 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 $V_k := \cup_{x \in F} B\left(x, 1/k\right)$, where $B\left(x, \delta\right) := \left\{y \in \mathbb{R}^n : |y - x| < \delta\right\}$. 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 5.10.
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 $\mu\left(B \setminus K\right) < \varepsilon$. **Hint:** take $K := F \cap \left\{x \in \mathbb{R}^n : |x| \leq n\right\}$ for some sufficiently large n.

5.4 Construction of Measures

Remark 5.14. Let us recall from Proposition 4.3 and Remark 4.4 that a finitely additive measure $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure on $\mathcal{A}$ iff $\mu\left(A_n\right) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Furthermore if $\mu\left(X\right) < \infty$, 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 5.15. *Let μ be a premeasure on an algebra $\mathcal{A}$, then μ has a unique extension (still called μ) to a function on $\mathcal{A}_\sigma$ satisfying the following properties.*

1. ***(Continuity)*** *If $A_n \in \mathcal{A}$ and $A_n \uparrow A \in \mathcal{A}_\sigma$, then $\mu\left(A_n\right) \uparrow \mu\left(A\right)$ as $n \to \infty$.*
2. ***(Monotonicity)*** *If $A, B \in \mathcal{A}_\sigma$ with $A \subset B$ then $\mu\left(A\right) \leq \mu\left(B\right)$.*
3. ***(Strong Additivity)*** *If $A, B \in \mathcal{A}_\sigma$, then*
$$\mu\left(A \cup B\right) + \mu\left(A \cap B\right) = \mu\left(A\right) + \mu\left(B\right). \tag{5.3}$$
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*
$$\mu\left(\cup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu\left(A_n\right). \tag{5.4}$$

*5. (σ - Additivity **on** $\mathcal{A}_\sigma$) The function μ is countably additive on $\mathcal{A}_\sigma$.*

Proof. Let A, B be sets in $\mathcal{A}_\sigma$ such that $A \subset B$ and suppose $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Since $B_m \cap A_n \uparrow A_n$ as $m \to \infty$, the continuity of μ on $\mathcal{A}$ implies,

$$\mu(A_n) = \lim_{m\to\infty} \mu(B_m \cap A_n) \le \lim_{m\to\infty} \mu(B_m).$$

We may let $n \to \infty$ in this inequality to find,

$$\lim_{n\to\infty} \mu(A_n) \le \lim_{m\to\infty} \mu(B_m). \tag{5.5}$$

Using this equation when $B = A$, implies, $\lim_{n\to\infty} \mu(A_n) = \lim_{m\to\infty} \mu(B_m)$ whenever $A_n \uparrow A$ and $B_n \uparrow A$. Therefore it is unambiguous to define $\mu(A)$ by;

$$\mu(A) = \lim_{n\to\infty} \mu(A_n)$$

for any sequence $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A$. With this definition, the continuity of μ is clear and the monotonicity of μ follows from Eq. (5.5).

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 $\mathcal{A}_\sigma$.

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,

$$\mu\left(\cup_{n=1}^N A_{n,N}\right) \le \sum_{n=1}^N \mu(A_{n,N}) \le \sum_{n=1}^N \mu(A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.6}$$

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. (5.6) to conclude Eq. (5.4) holds.

If we further assume that $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$ is a disjoint sequence, by the finite additivity and monotonicity of μ on $\mathcal{A}_\sigma$, we have

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

The previous two inequalities show μ is σ – additive on $\mathcal{A}_\sigma$. ∎

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

$$\mu(A) := \mu(X) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{5.7}$$

Lemma 5.16. *Suppose μ is a finite premeasure on an algebra, $\mathcal{A} \subset 2^X$, and μ has been extended to $\mathcal{A}_\delta \cup \mathcal{A}_\sigma$ as described in Proposition 5.15 and Eq. (5.7) above.*

1. *If $A \in \mathcal{A}_\delta$ and $A_n \in \mathcal{A}$ such that $A_n \downarrow A$, then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*
2. *μ is additive when restricted to $\mathcal{A}_\delta$.*
3. *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 $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 5.15 it follows that
$$\mu(A) = \mu(X) - \mu(A^c) = \mu(X) - \lim_{n\to\infty} \mu(A_n^c)$$
$$= \lim_{n\to\infty} [\mu(X) - \mu(A_n^c)] = \lim_{n\to\infty} \mu(A_n).$$
2. Suppose $A, B \in \mathcal{A}_\delta$ are disjoint sets 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 therefore,
$$\mu(A \cup B) = \lim_{n\to\infty} \mu(A_n \cup B_n) = \lim_{n\to\infty} [\mu(A_n) + \mu(B_n) - \mu(A_n \cap B_n)]$$
$$= \mu(A) + \mu(B)$$
wherein the last equality we have used Proposition 4.3.
3. By assumption, $X = A^c \cup C$. So applying the strong additivity of μ on $\mathcal{A}_\sigma$ in Eq. (5.3) with $A \to A^c \in \mathcal{A}_\sigma$ and $B \to C \in \mathcal{A}_\sigma$ shows
$$\mu(X) + \mu(C \setminus A) = \mu(A^c \cup C) + \mu(A^c \cap C)$$
$$= \mu(A^c) + \mu(C) = \mu(X) - \mu(A) + \mu(C).$$

■

Definition 5.17 (Measurable Sets). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^X$. We say that $B \subset X$ is* ***measurable*** *if 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 $\mu(C \setminus A) < \varepsilon$. We will denote the collection of measurable subsets of X by $\mathcal{B} = \mathcal{B}(\mu)$. We also define $\bar{\mu} : \mathcal{B} \to [0, \mu(X)]$ by*
$$\bar{\mu}(B) = \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\}. \tag{5.8}$$

Remark 5.18. If $B \in \mathcal{B}$, $\varepsilon > 0$, $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ are such that $A \subset B \subset C$ and $\mu(C \setminus A) < \varepsilon$, then $\mu(A) \le \bar{\mu}(B) \le \mu(C)$ and in particular,
$$0 \le \bar{\mu}(B) - \mu(A) < \varepsilon, \text{ and } 0 \le \mu(C) - \bar{\mu}(B) < \varepsilon. \tag{5.9}$$

Indeed, if $C' \in \mathcal{A}_\sigma$ with $B \subset C'$, then $A \subset C'$ and so by Lemma 5.16,
$$\mu(A) \le \mu(C' \setminus A) + \mu(A) = \mu(C')$$
from which it follows that $\mu(A) \le \bar{\mu}(B)$. The fact that $\bar{\mu}(B) \le \mu(C)$ follows directly from Eq. (5.8).

Theorem 5.19 (Finite Premeasure Extension Theorem). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^X$. Then $\mathcal{B}$ is a σ – algebra on X 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

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

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 5.16 to show,

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

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

$$\sum_{i=1}^\infty \mu(C_i) \leq \mu(X) + \varepsilon < \infty. \tag{5.10}$$

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

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

Therefore, using the sub-additivity of μ on $\mathcal{A}_\sigma$ and the estimate (5.10),

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

Since $\varepsilon > 0$ is arbitrary, it follows that $B \in \mathcal{B}$. Moreover by repeated use of Remark 5.18, we find

$$|\bar{\mu}(B) - \mu(A^n)| < \varepsilon + \sum_{i=n+1}^{\infty} \mu(C_i) \text{ and}$$

$$\left|\sum_{i=1}^{n} \bar{\mu}(B_i) - \mu(A^n)\right| = \left|\sum_{i=1}^{n} [\bar{\mu}(B_i) - \mu(A_i)]\right| \leq \sum_{i=1}^{n} |\bar{\mu}(B_i) - \mu(A_i)| \leq \varepsilon \sum_{i=1}^{n} 2^{-i} < \varepsilon.$$

Combining these estimates shows

$$\left|\bar{\mu}(B) - \sum_{i=1}^{n} \bar{\mu}(B_i)\right| < 2\varepsilon + \sum_{i=n+1}^{\infty} \mu(C_i)$$

which upon letting $n \to \infty$ gives,

$$\left|\bar{\mu}(B) - \sum_{i=1}^{\infty} \bar{\mu}(B_i)\right| \leq 2\varepsilon.$$

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 $\mathcal{B}$. ■

Theorem 5.20. *Suppose that μ is a σ – finite premeasure on an algebra $\mathcal{A}$. Then*

$$\bar{\mu}(B) := \inf\{\mu(C) : B \subset C \in \mathcal{A}_\sigma\} \;\; \forall \; B \in \sigma(\mathcal{A}) \tag{5.11}$$

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

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

Each μ_n is a premeasure (as is easily verified) on $\mathcal{A}$ and hence by Theorem 5.19 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. (5.11) holds. Let $B \in \sigma(\mathcal{A})$, $B_m = X_m \cap B$ and $\varepsilon > 0$ be given. By Theorem 5.19, there exists $C_m \in \mathcal{A}_\sigma$ such that $B_m \subset C_m \subset X_m$ 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

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

Thus

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

which, since $\varepsilon > 0$ is arbitrary, shows $\bar{\mu}$ satisfies Eq. (5.11). The uniqueness of the extension $\bar{\mu}$ is proved in Exercise 5.1. ■

Example 5.21. 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 5.22. *Lebesgue measure m is invariant under translations, i.e. for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$,*

$$m(x + B) = m(B). \tag{5.12}$$

Moreover, m is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0,1]) = 1$ and Eq. (5.12) holds for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B) = |\lambda| \, m(B) \tag{5.13}$$

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 5.1. 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. (5.13) 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

$$m_\lambda((a,b]) = \lambda^{-1} m((\lambda a, \lambda b]) = \lambda^{-1}(\lambda b - \lambda a) = b - a$$

if $\lambda > 0$ and

$$m_\lambda((a,b]) = |\lambda|^{-1} m([\lambda b, \lambda a)) = -|\lambda|^{-1} (\lambda b - \lambda a) = b - a$$

if $\lambda < 0$. Hence $m_\lambda = m$. ∎

5.5 Completions of Measure Spaces

Definition 5.23. *A set $E \subset X$ 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 $x \in X$, we will use the terminology P a.e. (to be read P almost everywhere) to mean*

$$E := \{x \in X : P \text{ is false for } x\}$$

is a null set. For example if f and g are two measurable functions on $(X, \mathcal{B}, \mu)$, $f = g$ a.e. means that $\mu(f \neq g) = 0$.

Definition 5.24. *A measure space $(X, \mathcal{B}, \mu)$ is* ***complete*** *if every subset of a null set is in $\mathcal{B}$, i.e. for all $F \subset X$ such that $F \subset E \in \mathcal{B}$ with $\mu(E) = 0$ implies that $F \in \mathcal{B}$.*

Proposition 5.25 (Completion of a Measure). *Let $(X, \mathcal{B}, \mu)$ be a measure space. Set*

$$\begin{aligned} \mathcal{N} = \mathcal{N}^\mu &:= \{N \subset X : \exists\, F \in \mathcal{B} \text{ such that } N \subset F \text{ and } \mu(F) = 0\}, \\ \mathcal{B} = \bar{\mathcal{B}}^\mu &:= \{A \cup N : A \in \mathcal{B} \text{ and } N \in \mathcal{N}\} \text{ and} \\ \bar{\mu}(A \cup N) &:= \mu(A) \text{ for } A \in \mathcal{B} \text{ and } N \in \mathcal{N}, \end{aligned}$$

see Fig. 5.1. 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 $(X, \bar{\mathcal{B}}, \bar{\mu})$ 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 $X, \emptyset \in \bar{\mathcal{B}}$. Let $A \in \mathcal{B}$ and $N \in \mathcal{N}$ and choose $F \in \mathcal{B}$ such

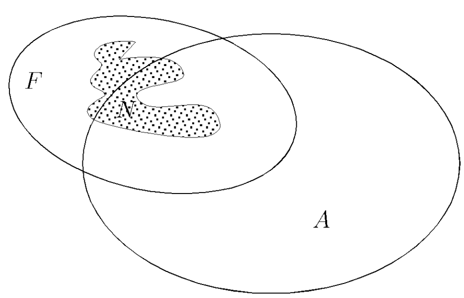


Fig. 5.1. Completing a σ – algebra.

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 $\sigma-$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. ∎

5.6 A Baby Version of Kolmogorov's Extension Theorem

For this section, let Λ be a finite set, $\Omega := \Lambda^\infty := \Lambda^{\mathbb{N}}$, and let $\mathcal{A}$ denote the collection of **cylinder subsets of** Ω, where $A \subset \Omega$ is a **cylinder set** iff there exists $n \in \mathbb{N}$ and $B \subset \Lambda^n$ such that

$$A = B \times \Lambda^\infty := \{\omega \in \Omega : (\omega_1, \dots, \omega_n) \in B\}.$$

Observe that we may also write A as $A = B' \times \Lambda^\infty$ where $B' = B \times \Lambda^k \subset \Lambda^{n+k}$ for any $k \geq 0$.

Exercise 5.3. Show $\mathcal{A}$ is an algebra.

Lemma 5.26. *Suppose $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

Proof. Since $A_n \in \mathcal{A}$, we may find $N_n \in \mathbb{N}$ and $B_n \subset \Lambda^{N_n}$ such that $A_n = B_n \times \Lambda^\infty$. Using the observation just prior to this Lemma, we may assume that $\{N_n\}_{n=1}^\infty$ is a strictly increasing sequence.

By assumption, there exists $\omega(n) = (\omega_1(n), \omega_2(n), \dots) \in \Omega$ such that $\omega(n) \in A_n$ for all n. Moreover, since $\omega(n) \in A_n \subset A_k$ for all $k \leq n$, it follows that

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } k \leq n. \tag{5.14}$$

Since Λ is a finite set, we may find a $\lambda_1 \in \Lambda$ and an infinite subset, $\Gamma_1 \subset \mathbb{N}$ such that $\omega_1(n) = \lambda_1$ for all $n \in \Gamma_1$. Similarly, there exists $\lambda_2 \in \Lambda$ and an infinite set, $\Gamma_2 \subset \Gamma_1$, such that $\omega_2(n) = \lambda_2$ 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 $\lambda_j \in \Lambda$ such that $\Gamma_1 \supset \Gamma_2 \supset \Gamma_3 \supset \dots$ and $\omega_j(n) = \lambda_j$ for all $n \in \Gamma_j$.

We are now going to complete the proof by showing that $\lambda := (\lambda_1, \lambda_2, \dots)$ is in $\cap_{n=1}^\infty A_n$. By the construction above, for all $N \in \mathbb{N}$ we have

$$(\omega_1(n), \dots, \omega_N(n)) = (\lambda_1, \dots, \lambda_N) \text{ for all } n \in \Gamma_N.$$

Taking $N = N_k$ and $n \in \Gamma_{N_k}$ with $n \geq k$, we learn from Eq. (5.14) that

$$(\lambda_1, \dots, \lambda_{N_k}) = (\omega_1(n), \dots, \omega_{N_k}(n)) \in B_k.$$

But this is equivalent to showing $\lambda \in A_k$. Since $k \in \mathbb{N}$ was arbitrary it follows that $\lambda \in \cap_{n=1}^{\infty} A_n$. ■

Theorem 5.27 (Kolmogorov's Extension Theorem I.). *Continuing the notation above, every finitely additive probability measure,* $P : \mathcal{A} \to [0,1]$, *has a unique extension to a probability measure on* $\sigma(\mathcal{A})$.

Proof. From Theorem 5.19, 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 5.26, if $A_n \in \mathcal{A}$ 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 $\lim_{n\to\infty} P(A_n) = 0$. ■

Given a probability measure, $P : \sigma(\mathcal{A}) \to [0,1]$ and $n \in \mathbb{N}$ and $(\lambda_1, \dots, \lambda_n) \in \Lambda^n$, let

$$p_n(\lambda_1, \dots, \lambda_n) := P(\{\omega \in \Omega : \omega_1 = \lambda_1, \dots, \omega_n = \lambda_n\}). \tag{5.15}$$

Exercise 5.4 (Consistency Conditions). If p_n is defined as above, show:

1. $\sum_{\lambda\in\Lambda} p_1(\lambda) = 1$ and
2. for all $n \in \mathbb{N}$ and $(\lambda_1, \dots, \lambda_n) \in \Lambda^n$,

$$p_n(\lambda_1, \dots, \lambda_n) = \sum_{\lambda\in\Lambda} p_{n+1}(\lambda_1, \dots, \lambda_n, \lambda).$$

Exercise 5.5 (Converse to 5.4). Suppose for each $n \in \mathbb{N}$ we are given functions, $p_n : \Lambda^n \to [0,1]$ such that the consistency conditions in Exercise 5.4 hold. Then there exists a unique probability measure, P on $\sigma(\mathcal{A})$ such that Eq. (5.15) holds for all $n \in \mathbb{N}$ and $(\lambda_1, \dots, \lambda_n) \in \Lambda^n$.

Example 5.28 (Existence of iid simple R.V.s). Suppose now that $q : \Lambda \to [0,1]$ is a function such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$. Then there exists a unique probability measure P on $\sigma(\mathcal{A})$ such that, for all $n \in \mathbb{N}$ and $(\lambda_1, \dots, \lambda_n) \in \Lambda^n$, we have

$$P(\{\omega \in \Omega : \omega_1 = \lambda_1, \dots, \omega_n = \lambda_n\}) = q(\lambda_1) \dots q(\lambda_n).$$

This is a special case of Exercise 5.5 with $p_n(\lambda_1, \dots, \lambda_n) := q(\lambda_1) \dots q(\lambda_n)$.

6

Random Variables

6.1 Measurable Functions

Definition 6.1. *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 6.7 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

Definition 6.2. *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 6.3. 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 6.4 (Characteristic 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 6.5. Suppose $f : X \to Y$ with Y being a finite set and $\mathcal{F} = 2^{\Omega}$. Then f is measurable iff $f^{-1}(\{y\}) \in \mathcal{M}$ for all $y \in Y$.

Proposition 6.6. *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, using Lemma 3.26,

$$f^{-1}(\mathcal{F}) = f^{-1}(\sigma(\mathcal{E})) = \sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}.$$

■

Corollary 6.7. *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 6.1. Prove Corollary 6.7. **Hint:** See Exercise 3.9.

Exercise 6.2. 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 6.3. 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 $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \mathbf{in}\ \mathbb{R}\} \in \mathcal{M}$.

Exercise 6.4. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{R}})$ – measurable.

Definition 6.8. *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 6.9 (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 measurable.*
2. *Suppose there exist $A_n \in \mathcal{M}$ such that $X = \cup_{n=1}^\infty A_n$ and $f|A_n$ is $\mathcal{M}_{A_n}$ 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 $f^{-1}(B) \in \mathcal{M}$. ■

The proof of the following exercise is routine and will be left to the reader.

Proposition 6.10. *Let $(X, \mathcal{M}, \mu)$ be a measure space, $(Y, \mathcal{F})$ be a measurable space and $f : X \to Y$ be a measurable map. Define a function $\nu : \mathcal{F} \to [0, \infty]$ by $\nu(A) := \mu(f^{-1}(A))$ for all $A \in \mathcal{F}$. Then ν is a measure on $(Y, \mathcal{F})$. (In the future we will denote ν by $f_*\mu$ or $\mu \circ f^{-1}$ and call $f_*\mu$ the **push-forward of** μ **by** f or the **law of** f **under** μ.*

Theorem 6.11. *Given a distribution function, $F : \mathbb{R} \to [0, 1]$ let $G : (0, 1) \to \mathbb{R}$ be defined (see Figure 6.1) 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 $-\infty < a < b < \infty$.

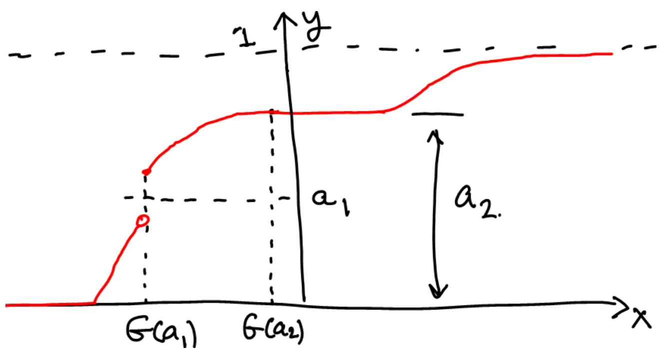


Fig. 6.1. A pictorial definition of G.

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

$$G^{-1}((0, x_0]) = \{y : G(y) \leq x_0\} = (0, F(x_0)] \cap \mathbb{R}, \tag{6.1}$$

see Figure 6.2.

To give a formal proof of Eq. (6.1), $G(y) = \inf \{x : F(x) \geq y\} \leq x_0$, 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

$$G(F(x_0)) = \inf \{x : F(x) \geq F(x_0)\} \leq x_0$$

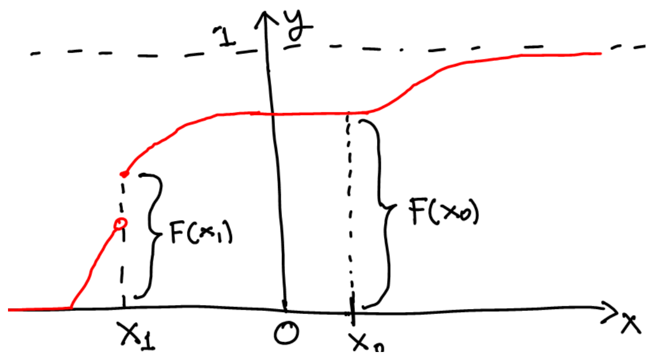


Fig. 6.2. As can be seen from this picture, $G(y) \leq x_0$ iff $y \leq F(x_0)$ and similalry, $G(y) \leq x_1$ iff $y \leq x_1$.

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 6.12 (Durret's Version). *Given a distribution function,* $F : \mathbb{R} \to [0,1]$ *let* $Y : (0,1) \to \mathbb{R}$ *be defined (see Figure 6.3) 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* $-\infty < a < b < \infty$.

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

$$F(Y(x)-) \leq x \leq F(Y(x)).$$

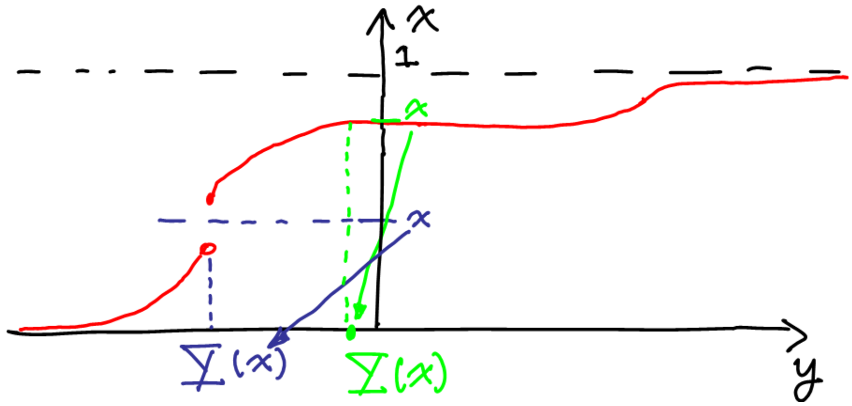


Fig. 6.3. A pictorial definition of $Y(x)$.

We will now show

$$\{x \in (0,1) : Y(x) \leq y_0\} = (0, F(y_0)] \cap (0,1). \tag{6.2}$$

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. (6.2), 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. ∎

Lemma 6.13 (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 6.14 (σ – Algebras Generated by Functions). *Let X be a set and suppose there is a collection of measurable spaces $\{(Y_\alpha, \mathcal{F}_\alpha) : \alpha \in A\}$ and functions $f_\alpha : X \to Y_\alpha$ for all $\alpha \in A$. Let $\sigma(f_\alpha : \alpha \in A)$ denote the smallest σ – algebra on X such that each f_α is measurable, i.e.*

$$\sigma(f_\alpha : \alpha \in A) = \sigma(\cup_\alpha f_\alpha^{-1}(\mathcal{F}_\alpha)).$$

Example 6.15. 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 6.16. *Assuming the notation in Definition 6.14 and additionally let $(Z, \mathcal{M})$ be a measurable space and $g : Z \to X$ be a function. Then g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in A))$ – measurable iff $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$–measurable for all $\alpha \in A$.*

Proof. ($\Rightarrow$) If g is $(\mathcal{M}, \sigma(f_\alpha : \alpha \in A))$ – measurable, then the composition $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable by Lemma 6.13. ($\Leftarrow$) Let

$$\mathcal{G} = \sigma(f_\alpha : \alpha \in A) = \sigma\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right).$$

If $f_\alpha \circ g$ is $(\mathcal{M}, \mathcal{F}_\alpha)$ – measurable for all α, then

$$g^{-1} f_\alpha^{-1}(\mathcal{F}_\alpha) \subset \mathcal{M} \, \forall \, \alpha \in A$$

and therefore

$$g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right) = \cup_{\alpha \in A} g^{-1} f_\alpha^{-1}(\mathcal{F}_\alpha) \subset \mathcal{M}.$$

Hence

$$g^{-1}(\mathcal{G}) = g^{-1}\left(\sigma\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right)\right) = \sigma(g^{-1}\left(\cup_{\alpha \in A} f_\alpha^{-1}(\mathcal{F}_\alpha)\right) \subset \mathcal{M}$$

which shows that g is $(\mathcal{M}, \mathcal{G})$ – measurable. ■

Definition 6.17. *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 6.18. *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 3.26 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 6.19. For $i = 1, 2, \dots, n$, 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 6.20. *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 $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \dots, \pi_n)$ 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 $i = 1, 2, \dots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$.*

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 $i = 1, 2, \ldots, n$ 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 π_i is $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable, it follows that $\sigma(\pi_1, \ldots, \pi_n) \subset \mathcal{B}_{\mathbb{R}^n}$. Moreover, if (a,b) is as in Eq. (6.3), then

$$(a,b) = \cap_{i=1}^n \pi_i^{-1}((a_i, b_i)) \in \sigma(\pi_1, \ldots, \pi_n).$$

Therefore, $\mathcal{E} \subset \sigma(\pi_1, \ldots, \pi_n)$ and $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E}) \subset \sigma(\pi_1, \ldots, \pi_n)$.

Assertion 3. If $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$, then

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

■

Corollary 6.21. *If $(X, \mathcal{M})$ is a measurable space, then*

$$f = (f_1, f_2, \ldots, f_n) : X \to \mathbb{R}^n$$

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 6.20 and Proposition 6.16. ■

Corollary 6.22. *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 6.23. *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

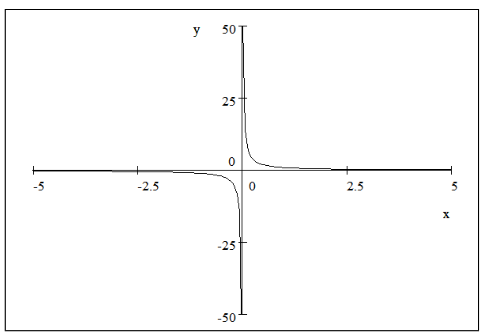
$$i(z) = \begin{cases} \frac{1}{z} & \text{if} \quad z \neq 0 \\ 0 & \text{if} \quad z = 0. \end{cases}$$

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 6.24. For the real case of Lemma 6.23, define i as above but now take z to real. From the plot of i, Figure 6.24, the reader may easily verify that $i^{-1}((-\infty, a])$ is an infinite half interval for all a and therefore i is measurable. $\frac{1}{x}$



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 6.25 (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,

$$i^{-1}(\mathcal{B}_{\bar{\mathbb{R}}}) = \sigma\left(i^{-1}\left(\left\{[a,\infty] : a \in \bar{\mathbb{R}}\right\}\right)\right) = \sigma\left(\left\{i^{-1}([a,\infty]) : a \in \bar{\mathbb{R}}\right\}\right)$$
$$= \sigma\left(\left\{[a,\infty] \cap \mathbb{R} : a \in \bar{\mathbb{R}}\right\}\right) = \sigma\left(\{[a,\infty) : a \in \mathbb{R}\}\right) = \mathcal{B}_{\mathbb{R}}.$$

Thus we have shown

$$\mathcal{B}_{\mathbb{R}} = i^{-1}(\mathcal{B}_{\bar{\mathbb{R}}}) = \{A \cap \mathbb{R} : A \in \mathcal{B}_{\bar{\mathbb{R}}}\}.$$

This implies:

1. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \implies A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ 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 6.5 and 6.6.

Corollary 6.26. *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*

$$f^0(x) := 1_{\mathbb{R}}(f(x)) = \begin{cases} f(x) & \text{if } \ f(x) \in \mathbb{R} \\ 0 & \text{if } f(x) \in \{\pm\infty\} \end{cases}$$

is measurable.

Corollary 6.27. *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 6.5. Prove Corollary 6.26 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 6.7. Use Proposition 6.25 to handle item 4.

Exercise 6.6. Prove Corollary 6.27.

Proposition 6.28 (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. ■

Definition 6.29. *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 6.30. *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 6.28 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ■

Definition 6.31. *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 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 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

$$\varphi = \sum_{y\in\mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{6.7}$$

The next theorem shows that simple functions are "pointwise dense" in the space of measurable functions.

Theorem 6.32 (Approximation Theorem). *Let $f : X \to [0, \infty]$ be measurable and define, see Figure 6.4,*

$$\varphi_n(x) := \sum_{k=0}^{n2^n-1} \frac{k}{2^n} 1_{f^{-1}\left(\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right)}(x) + n1_{f^{-1}((n2^n,\infty])}(x)$$
$$= \sum_{k=0}^{n2^n-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}}(x) + n1_{\{f>n2^n\}}(x)$$

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 $n \to \infty$.

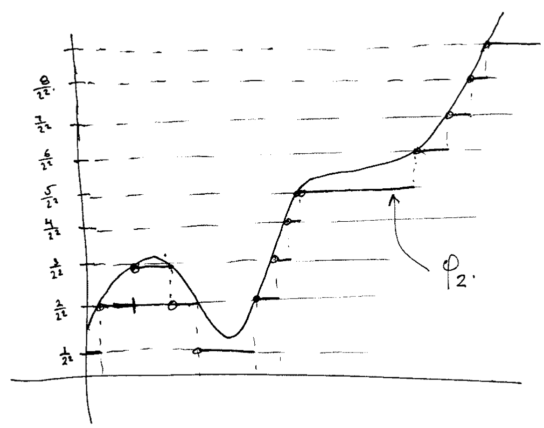


Fig. 6.4. Constructing simple functions approximating a function, $f : X \to [0, \infty]$.

Proof. Since

$$\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] = \left(\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right] \cup \left(\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right],$$

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 $x \in f^{-1}((2^n, 2^{n+1}])$, $\varphi_{n+1}(x) \geq 2^n = \varphi_n(x)$. Therefore $\varphi_n \leq \varphi_{n+1}$ for all n. It is clear by construction that $\varphi_n(x) \leq f(x)$ for all x and that $0 \leq f(x) - \varphi_n(x) \leq 2^{-n}$ if $x \in X_{2^n}$. 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 simple functions such that $\varphi_n^{\pm} \uparrow f_{\pm}$ as $n \to \infty$ and define $\varphi_n = \varphi_n^+ - \varphi_n^-$. Then

$$|\varphi_n| = \varphi_n^+ + \varphi_n^- \leq \varphi_{n+1}^+ + \varphi_{n+1}^- = |\varphi_{n+1}|$$

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 $n \to \infty$. ∎

6.2 Factoring Random Variables

Lemma 6.33. *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 $H = \sum a_i 1_{A_i}$ 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 $\bar{\mathbb{R}}$. ∎

The following is an immediate corollary of Proposition 6.16 and Lemma 6.33.

Corollary 6.34. *Let X and A be sets, and suppose for $\alpha \in A$ we are give a measurable space $(Y_\alpha, \mathcal{F}_\alpha)$ and a function $f_\alpha : X \to Y_\alpha$. Let $Y := \prod_{\alpha \in A} Y_\alpha$, $\mathcal{F} := \otimes_{\alpha \in A} \mathcal{F}_\alpha$ be the product σ – algebra on Y and $\mathcal{M} := \sigma(f_\alpha : \alpha \in A)$*

be the smallest σ – algebra on X such that each f_α is measurable. Then the function $F : X \to Y$ defined by $[F(x)]_\alpha := f_\alpha(x)$ for each $\alpha \in A$ is $(\mathcal{M}, \mathcal{F})$ – measurable and a function $H : X \to \bar{\mathbb{R}}$ is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable iff there exists a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function h from Y to $\bar{\mathbb{R}}$ such that $H = h \circ F$.

7

Independence

7.1 $\pi - \lambda$ and Monotone Class Theorems

Definition 7.1. *Let $\mathcal{C} \subset 2^X$ be a collection of sets.*

1. *$\mathcal{C}$ is a **monotone class** if it is closed under countable increasing unions and countable decreasing intersections,*
2. *$\mathcal{C}$ is a π – **class** if it is closed under finite intersections and*
3. *$\mathcal{C}$ is a λ–**class** if $\mathcal{C}$ satisfies the following properties:*
   a) *$X \in \mathcal{C}$*
   b) *If $A, B \in \mathcal{C}$ and $A \subset B$, then $B \setminus A \in \mathcal{C}$. (Closed under proper differences.)*
   c) *If $A_n \in \mathcal{C}$ and $A_n \uparrow A$, then $A \in \mathcal{C}$. (Closed under countable increasing unions.)*

Remark 7.2. If $\mathcal{C}$ is a collection of subsets of Ω which is both a λ – class and a π – system then $\mathcal{C}$ is a σ – algebra. Indeed, since $A^c = X \setminus A$, we see that any λ - system is closed under complementation. If $\mathcal{C}$ is also a π – system, it is closed under intersections and therefore $\mathcal{C}$ is an algebra. Since $\mathcal{C}$ is also closed under increasing unions, $\mathcal{C}$ is a σ – algebra.

Lemma 7.3 (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. *$X \in \mathcal{L}$*
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 compliments 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 7.4 (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. $X \in \mathcal{L}^C$, b. if $A \subset B$ with $A, B \in \mathcal{L}^C$, then $A \cap C$, $B \cap C \in \mathcal{L}$ with $A \cap C \subset B \setminus C$ and

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

Therefore $\mathcal{L}^C$ is closed under proper differences. Finally, 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 with out 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 7.2, $\mathcal{L}$ 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 7.5. *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

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

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. ∎

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

Example 7.6. 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 7.1. Suppose that μ and ν are two measure 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 (7.1). 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 7.5, $\mu_n = \nu_n$ 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 7.7. *A probability measure, P, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is uniquely determined by its distribution function,*

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

Definition 7.8. *Suppose that $\{X_i\}_{i=1}^n$ is a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$. The measure, $\mu = P \circ (X_1, \dots, X_n)^{-1}$ on $\mathcal{B}_{\mathbb{R}^n}$ is called the **joint distribution** 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 7.9. *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 7.5 with $\mathcal{P}$ being the π – systems defined by

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

for the first case and

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

for the second case. ■

Definition 7.10. *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 $(X, \mathcal{F}, Q)$ respectively. We write $(X_1, \ldots, X_n) \stackrel{d}{=} (Y_1, \ldots, Y_n)$ if $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ have the **same distribution**, i.e. if*

$$P((X_1, \ldots, X_n) \in B) = Q((Y_1, \ldots, Y_n) \in B) \text{ for all } B \in \mathcal{B}_{\mathbb{R}^n}.$$

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 $(X, \mathcal{F}, Q)$ we write $\{X_i\}_{i=1}^\infty \stackrel{d}{=} \{Y_i\}_{i=1}^\infty$ iff $(X_1, \ldots, X_n) \stackrel{d}{=} (Y_1, \ldots, Y_n)$ for all $n \in \mathbb{N}$.

Exercise 7.2. Let $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ be two sequences of random variables such that $\{X_i\}_{i=1}^\infty \stackrel{\mathrm{d}}{=} \{Y_i\}_{i=1}^\infty$. Let $\{S_n\}_{n=1}^\infty$ and $\{T_n\}_{n=1}^\infty$ be defined by, $S_n := X_1 + \cdots + X_n$ and $T_n := Y_1 + \cdots + Y_n$. Prove the following assertions.

1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1, \ldots, X_n) \stackrel{\mathrm{d}}{=} f(Y_1, \ldots, Y_n)$.
2. Use your result in item 1. to show $\{S_n\}_{n=1}^\infty \stackrel{\mathrm{d}}{=} \{T_n\}_{n=1}^\infty$.
   Hint: apply item 1. with $k = n$ and a judiciously chosen function, $f : \mathbb{R}^n \to \mathbb{R}^n$.
3. Show $\limsup_{n\to\infty} X_n \stackrel{\mathrm{d}}{=} \limsup_{n\to\infty} Y_n$ and similarly that $\liminf_{n\to\infty} X_n \stackrel{\mathrm{d}}{=} \liminf_{n\to\infty} Y_n$.
   Hint: with the aid of the set identity,
   $$\left\{\limsup_{n\to\infty} X_n \geq x\right\} = \{X_n \geq x \text{ i.o.}\},$$
   show
   $$P\left(\limsup_{n\to\infty} X_n \geq x\right) = \lim_{n\to\infty} \lim_{m\to\infty} P\left(\cup_{k=n}^m \{X_k \geq x\}\right).$$
   To use this identity you will also need to find $B \in \mathcal{B}_{\mathbb{R}^m}$ such that
   $$\cup_{k=n}^m \{X_k \geq x\} = \{(X_1, \ldots, X_m) \in B\}.$$

7.1.1 The Monotone Class Theorem

This subsection may be safely skipped!

Lemma 7.11 (Monotone Class Theorem*). *Suppose $\mathcal{A} \subset 2^X$ 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 $X \in \mathcal{A} \subset \mathcal{C}$) and finite intersections and increasing unions from which it easily follows that $\mathcal{C}$ is a $\sigma -$ algebra. ■

Exercise 7.3. Suppose that $\mathcal{A} \subset 2^\Omega$ is an algebra, $\mathcal{B} := \sigma(\mathcal{A})$, and P is a probability measure on $\mathcal{B}$. Show, using the $\pi - \lambda$ theorem, that for every $B \in \mathcal{B}$ there exists $A \in \mathcal{A}$ such that that $P(A \triangle B) < \varepsilon$. Here

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

is the symmetric difference of A and B.

Hints:

1. It may be useful to observe that

$$1_{A \triangle B} = |1_A - 1_B|$$

so that $P(A \triangle B) = \mathbb{E}|1_A - 1_B|$.

2. Also observe that if $B = \cup B_i$ and $A = \cup_i A_i$, then
$$B \setminus A \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and}$$
$$A \setminus B \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i$$
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
$$P(B \triangle A_n) = P(B \setminus A_n) + P(A_n \setminus B) \to P(B \setminus A) + P(A \setminus B) = P(A \triangle B).$$
5. Let $\mathcal{L}$ be the collection of sets B for which the assertion of the theorem holds. Show $\mathcal{L}$ is a λ – system which contains $\mathcal{A}$.

Solution to Exercise (7.3). 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 $P(B_i \triangle A_i) = \mathbb{E}|1_{A_i} - 1_{B_i}| < \varepsilon/2$ and therefore,
$$\begin{aligned}P((B_2 \setminus B_1) \triangle (A_2 \setminus A_1)) &\le P((A_1 \triangle B_1) \cup (A_2 \triangle B_2)) \\ &\le P((A_1 \triangle B_1)) + P((A_2 \triangle B_2)) < \varepsilon.\end{aligned}$$
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,
$$P([\cup_n B_n] \triangle [\cup_n A_n]) \le \sum_{n=1}^{\infty} P(B_n \triangle A_n) < \varepsilon.$$
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 $P\left(B \triangle A^N\right) < \varepsilon$.

7.2 Basic Properties of Independence

For this section we will suppose that $(\Omega, \mathcal{B}, P)$ is a probability space.

Definition 7.12. *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 $J \subset \{1, 2, \ldots, n\}$.

Observe that if $\{\mathcal{C}_i\}_{i=1}^n$, are independent classes then so are $\{\mathcal{C}_i \cup \{X\}\}_{i=1}^n$. Moreover, if we assume that $X \in \mathcal{C}_i$ 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 7.13. *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 with out loss of generality that $X \in \mathcal{C}_i$. Fix, $A_j \in \mathcal{C}_j$ for $j = 2, 3, \ldots, n$. We will begin by showing that

$$P(A \cap A_2 \cap \cdots \cap A_n) = P(A) P(A_2) \ldots P(A_n) \text{ for all } A \in \sigma(\mathcal{C}_1). \quad (7.1)$$

Since it is clear that this identity holds if $P(A_j) = 0$ for some $j = 2, \ldots, n$, we may assume that $P(A_j) > 0$ for $j \geq 2$. In this case we may define,

$$\begin{aligned} Q(A) &= \frac{P(A \cap A_2 \cap \cdots \cap A_n)}{P(A_2) \ldots P(A_n)} = \frac{P(A \cap A_2 \cap \cdots \cap A_n)}{P(A_2 \cap \cdots \cap A_n)} \\ &= P(A|A_2 \cap \cdots \cap A_n) \text{ for all } A \in \sigma(\mathcal{C}_1). \end{aligned}$$

Then equation Eq. (7.1) is equivalent to $P(A) = Q(A)$ on $\sigma(\mathcal{C}_1)$. But this is true by Proposition 7.5 using the fact that $Q = P$ on the π – system, $\mathcal{C}_1$.

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 7.14. *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 7.15. *If* $\{\mathcal{C}_t\}_{t\in T}$ *is a collection of independent classes such that each* $\mathcal{C}_t$ *is a* π *– system, then* $\{\sigma\left(\mathcal{C}_t\right)\}_{t\in T}$ *are independent as well.*

Example 7.16. Suppose that $\Omega=\Lambda^n$ where Λ is a finite set, $\mathcal{B}=2^\Omega$, $P\left(\{\omega\}\right)=\prod_{j=1}^n q_j\left(\omega_j\right)$ where $q_j:\Lambda\to[0,1]$ are functions such that $\sum_{\lambda\in\Lambda}q_j\left(\lambda\right)=1$. 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

$$\cap B_i=A_1\times A_2\times\cdots\times A_n$$

and we have

$$P\left(\cap B_i\right)=\sum_{\omega\in A_1\times A_2\times\cdots\times A_n}\prod_{i=1}^n q_i\left(\omega_i\right)=\prod_{i=1}^n\sum_{\lambda\in A_i}q_i\left(\lambda\right)$$

while

$$P\left(B_i\right)=\sum_{\omega\in\Lambda^{i-1}\times A_i\times\Lambda^{n-i}}\prod_{i=1}^n q_i\left(\omega_i\right)=\sum_{\lambda\in A_i}q_i\left(\lambda\right).$$

Definition 7.17. *A collections of random variables,* $\{X_t:t\in T\}$ *are* ***independent*** *iff* $\{\sigma\left(X_t\right):t\in T\}$ *are independent.*

Theorem 7.18. *Let* $\mathbb{X}:=\{X_t:t\in T\}$ *be a collection of random variables. Then the following are equivalent:*

1. *The collection* $\mathbb{X}$,
2. $$P\left(\cap_{t\in\Lambda}\{X_t\in A_t\}\right)=\prod_{t\in\Lambda}P\left(X_t\in A_t\right)$$
   for all finite subsets, $\Lambda\subset T$, *and all* $A_t\in\mathcal{B}_{\mathbb{R}}$ *for* $t\in\Lambda$.
3. $$P\left(\cap_{t\in\Lambda}\{X_t\le x_t\}\right)=\prod_{t\in\Lambda}P\left(X_t\le x_t\right)$$
   for all finite subsets, $\Lambda\subset T$, *and all* $x_t\in\mathbb{R}$ *for* $t\in\Lambda$.

Proof. The equivalence of 1. and 2. follows almost immediately form the definition of independence and the fact that $\sigma\left(X_t\right)=\{\{X_t\in A\}:A\in\mathcal{B}_{\mathbb{R}}\}$. Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 7.15 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 $\sigma\left(\mathcal{C}_t\right)=\sigma\left(X_t\right)$. ∎

Example 7.19. Continue the notation of Example 7.16 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 7.16.

Alternatively, from Exercise 4.1, we know that

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

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

$$\begin{aligned} P(X_1 \in A_1, \dots, X_n \in A_n) &= \mathbb{E}_P\left[\prod_{i=1}^n 1_{A_i}(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P\left[1_{A_i}(X_i)\right] \\ &= \prod_{i=1}^n P(X_i \in A_i) \end{aligned}$$

as desired.

Corollary 7.20. *A sequence of random variables, $\{X_j\}_{j=1}^k$ with countable ranges are independent iff*

$$P\left(\cap_{j=1}^k \{X_j = x_j\}\right) = \prod_{j=1}^k P(X_j = x_j) \tag{7.2}$$

for all $x_j \in \mathbb{R}$.

Proof. Observe that both sides of Eq. (7.2) are zero unless x_j is in the range of X_j for all j. Hence it suffices to verify Eq. (7.2) for those $x_j \in \mathrm{Ran}(X_j) =: R_j$ for all j. Now if $\{X_j\}_{j=1}^k$ are independent, then $\{X_j = x_j\} \in \sigma(X_j)$ for all $x_j \in \mathbb{R}$ and therefore Eq. (7.2) holds.

Conversely if Eq. (7.2) and $V_j \in \mathcal{B}_{\mathbb{R}}$, then

$$
\begin{aligned}
P\left(\cap_{j=1}^{k}\left\{X_j \in V_j\right\}\right) &= P\left(\cap_{j=1}^{k}\left[\sum_{x_j \in V_j \cap R_j}\left\{X_j = x_j\right\}\right]\right) \\
&= P\left(\sum_{(x_1,\dots,x_k) \in \prod_{j=1}^{k} V_j \cap R_j}\left[\cap_{j=1}^{k}\left\{X_j = x_j\right\}\right]\right) \\
&= \sum_{(x_1,\dots,x_k) \in \prod_{j=1}^{k} V_j \cap R_j} P\left(\left[\cap_{j=1}^{k}\left\{X_j = x_j\right\}\right]\right) \\
&= \sum_{(x_1,\dots,x_k) \in \prod_{j=1}^{k} V_j \cap R_j} \prod_{j=1}^{k} P\left(X_j = x_j\right) \\
&= \prod_{j=1}^{k} \sum_{x_j \in V_j \cap R_j} P\left(X_j = x_j\right) = \prod_{j=1}^{k} P\left(X_j \in V_j\right).
\end{aligned}
$$

■

Definition 7.21. *As 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*

$$P\left(X_n \in A\right) = P\left(X_k \in A\right) \text{ for all } k, n \in \mathbb{N} \text{ and } A \in \mathcal{B}_{\mathbb{R}}.$$

Observe that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables iff

$$P\left(X_1 \in A_1, \dots, X_n \in A_n\right) = \prod_{j=1}^{n} P\left(X_i \in A_i\right) = \prod_{j=1}^{n} P\left(X_1 \in A_i\right) = \prod_{j=1}^{n} \mu\left(A_i\right) \tag{7.3}$$

where $\mu = (X_1)_* P$. The identity in Eq. (7.3) is to hold for all $n \in \mathbb{N}$ and all $A_i \in \mathcal{B}_{\mathbb{R}}$.

Theorem 7.22 (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*

$$m\left(\left\{X_1 = i_1, \dots, X_k = i_i\right\}\right) = q_{i_1} \dots q_{i_k}$$

for all $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$ *and all* $k \in \mathbb{N}$. *(See Theorem 7.27 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

$$J_{i_1,i_2,\dots,i_k} := T_{i_k}\left(T_{i_{k-1}}\left(\dots T_{i_1}((0,1])\right)\right)$$

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 7.1. Repeated applications of Corollary 6.22 shows the functions, $X_k : (0,1] \to \mathbb{R}$ are measurable.

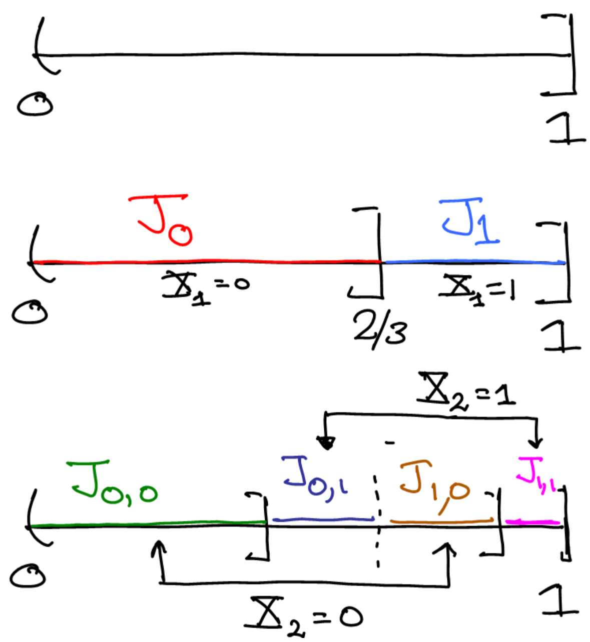


Fig. 7.1. Here we suppose that $p_0 = 2/3$ and $p_1 = 1/3$ and then we construct J_l and $J_{l,k}$ for $l, k \in \{0, 1\}$.

Observe that

$$m(T_i((a,b])) = q_i(b-a) = q_i m((a,b]), \tag{7.4}$$

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 7.23 (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 7.22 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 7.24. *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 7.20 to shows that random variables constructed in Example 5.28 or Theorem 7.22 fit the bill. ■

Proposition 7.25. *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 $P(X_n = 0 \text{ a.a.}) = P(X_n = 1 \text{ a.a.})$. 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

$$\left\{U < \frac{1}{2}\right\} = \{X_1 = 0\},$$
$$\left\{U < \frac{1}{4}\right\} = \{X_1 = 0, X_2 = 0\} \text{ and}$$
$$\left\{\frac{1}{2} \le U < \frac{3}{4}\right\} = \{X_1 = 1, X_2 = 0\}$$

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

$$P(U < x) = x. \tag{7.5}$$

The proof is by induction on n. Indeed, we have already verified (7.5) when $n = 1, 2$. Suppose we have verified (7.5) 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

$$P(U \le x) = (0 \vee x) \wedge 1.$$

■

Lemma 7.26. *Suppose that $\{\mathcal{B}_t : t \in T\}$ is an independent family of $\sigma-$ fields. And further assume that $T = \sum_{s\in S} T_s$ and let*

$$\mathcal{B}_{T_s} = \vee_{t\in T_s}\mathcal{B}_s = \sigma\left(\cup_{t\in T_s}\mathcal{B}_s\right).$$

Then $\{\mathcal{B}_{T_s}\}_{s\in S}$ is an independent family of σ fields.

Proof. Let

$$\mathcal{C}_s = \{\cap_{\alpha\in K} B_\alpha : B_\alpha \in \mathcal{B}_\alpha,\ K \subset\subset T_s\}.$$

It is now easily checked that $\{\mathcal{C}_s\}_{s\in S}$ is an independent family of $\pi-$systems. Therefore $\{\mathcal{B}_{T_s} = \sigma\left(\mathcal{C}_s\right)\}_{s\in S}$ is an independent family of $\sigma-$ algebras. ■

We may now show the existence of independent random variables with arbitrary distributions.

Theorem 7.27. *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 Law $(Y_n) := P \circ Y_n^{-1} = \mu_n$ for all n.*

Proof. By Theorem 7.24, 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 7.25, U_i is uniformly distributed on $[0, 1]$. Moreover by the grouping Lemma 7.26, $\left\{\sigma\left(\{X_{i,j}\}_{j=1}^\infty\right)\right\}_{i=1}^\infty$ are independent $\sigma-$ 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 6.11, $Y_i := G_i(U_i)$ 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. ■

7.2.1 An Example of Ranks

Let $\{X_n\}_{n=1}^\infty$ be i.i.d. with common continuous distribution function, F. In this case we have, for any $i \neq j$, that

$$P(X_i = X_j) = \mu_F \otimes \mu_F\left(\{(x, x) : x \in \mathbb{R}\}\right) = 0.$$

This may be proved directly with some work or will be an easy consequence of Fubini's theorem to be considered later, see Example 10.11 below. For the direct proof, let $\{a_l\}_{l=-\infty}^\infty$ be a sequence such that, $a_l < a_{l+1}$ for all $l \in \mathbb{Z}$, $\lim_{l\to\infty} a_l = \infty$ and $\lim_{l\to-\infty} a_l = -\infty$. Then

$$\{(x,x) : x \in \mathbb{R}\} \subset \cup_{l \in \mathbb{Z}} [(a_l, a_{l+1}] \times (a_l, a_{l+1}]]$$

and therefore,

$$\begin{aligned} P(X_i = X_j) &\le \sum_{l \in \mathbb{Z}} P(X_i \in (a_l, a_{l+1}], X_j \in (a_l, a_{l+1}]) = \sum_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)]^2 \\ &\le \sup_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)] \sum_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)] = \sup_{l \in \mathbb{Z}} [F(a_{l+1}) - F(a_l)]. \end{aligned}$$

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

$$P(X_i = X_j) \le \limsup_{N \to \infty} \sup_{l \in \mathbb{Z}} \left[F\left(\frac{l+1}{N}\right) - F\left(\frac{l}{N}\right) \right] = 0.$$

Let R_n denote the "rank" of X_n in the list $(X_1, \dots, X_n)$, i.e.

$$R_n := \sum_{j=1}^{n} 1_{X_j > X_n} = \#\{j \le n : X_j > X_n\}.$$

For example if $(X_1, X_2, X_3, X_4, X_5, \dots) = (9, -8, 3, 7, 23, \dots)$, we have $R_1 = 1$, $R_2 = 2$, $R_3 = 2$, and $R_4 = 2$, $R_5 = 1$. 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, $(X_2, X_3, X_4, X_1, X_5)$. As another example, if $R_i = i$ for $i = 1, 2, \dots, n$, then $X_n < X_{n-1} < \dots < X_1$.

Theorem 7.28 (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, $(X_1, \dots, X_n)$ 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,

$$\Omega = \sum_{\sigma} \{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}$$

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 $(\varepsilon_1, \ldots, \varepsilon_n)$, (here $\varepsilon_i \in \mathbb{N}$ with $\varepsilon_i \le i$) of $(R_1, \ldots, R_n)$ there is a permutation, $\sigma = \sigma(\varepsilon_1, \ldots, \varepsilon_n)$ such that $X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}$. From this it follows that

$$\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\} = \{X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}\}$$

and therefore,

$$P(\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\}) = P(X_{\sigma 1} < X_{\sigma 2} < \cdots < X_{\sigma n}) = \frac{1}{n!}.$$

Since

$$\begin{aligned} P(\{R_n = \varepsilon_n\}) &= \sum_{(\varepsilon_1, \ldots \varepsilon_{n-1})} P(\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\}) \\ &= \sum_{(\varepsilon_1, \ldots \varepsilon_{n-1})} \frac{1}{n!} = (n-1)! \cdot \frac{1}{n!} = \frac{1}{n} \end{aligned}$$

we have shown that

$$P(\{(R_1, \ldots, R_n) = (\varepsilon_1, \ldots, \varepsilon_n)\}) = \frac{1}{n!} = \prod_{j=1}^n \frac{1}{j} = \prod_{j=1}^n P(\{R_j = \varepsilon_j\}).$$

■

7.3 Borel-Cantelli Lemmas

Lemma 7.29 (First Borel Cantelli-Lemma). *Suppose that* $\{A_n\}_{n=1}^\infty$ *are measurable sets. If*

$$\sum_{n=1}^\infty P(A_n) < \infty, \tag{7.6}$$

then

$$P(\{A_n \text{ i.o.}\}) = 0.$$

Proof. First Proof. We have

$$P(\{A_n \text{ i.o.}\}) = P\left(\cap_{n=1}^{\infty} \cup_{k \geq n} A_k\right) = \lim_{n \to \infty} P\left(\cup_{k \geq n} A_k\right) \leq \lim_{n \to \infty} \sum_{k \geq n} P(A_k) = 0. \tag{7.7}$$

Second Proof. (**Warning:** this proof require integration theory which is developed below.) Equation (7.6) is equivalent to

$$\mathbb{E}\left[\sum_{n=1}^{\infty} 1_{A_n}\right] < \infty$$

from which it follows that

$$\sum_{n=1}^{\infty} 1_{A_n} < \infty \text{ a.s.}$$

which is equivalent to $P(\{A_n \text{ i.o.}\}) = 0$. ∎

Example 7.30. Suppose that $\{X_n\}$ are Bernoulli random variables with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. If

$$\sum p_n < \infty$$

then

$$P(X_n = 1 \text{ i.o.}) = 0$$

and hence

$$P(X_n = 0 \text{ a.a.}) = 1.$$

In particular,

$$P\left(\lim_{n \to \infty} X_n = 0\right) = 1.$$

Figure 7.2 below serves as motivation for the following elementary lemma on convex functions.

Lemma 7.31 (Convex Functions). *Suppose that* $\varphi \in PC^2((a,b) \to \mathbb{R})$[2] *with* $\varphi''(x) \geq 0$ *for almost all* $x \in (a,b)$. *Then* φ *satisfies;*

1. *for all* $x_0, x \in (a,b)$,

$$\varphi(x_0) + \varphi'(x_0)(x - x_0) \leq \varphi(x)$$

and

[2] 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 $j = 1, 2, \ldots, n$.

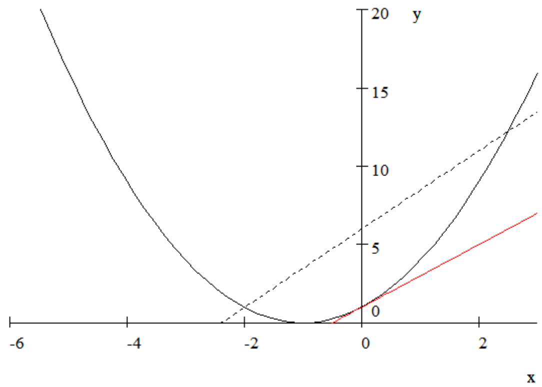


Fig. 7.2. A convex function, φ, along with a cord and a tangent line. Notice that the tangent line is always below φ and the cord lies above φ between the points of intersection of the cord with the graph of φ.

2. for all $u \leq v$ with $u, v \in (a, b)$,

$$\varphi(u + t(v - u)) \leq \varphi(u) + t(\varphi(v) - \varphi(u)) \;\; \forall\, t \in [0, 1].$$

(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 11.7 below for much more on convex functions.)

Proof. 1. Let

$$f(x) := \varphi(x) - [\varphi(x_0) + \varphi'(x_0)(x - x_0)].$$

Then $f(x_0) = f'(x_0) = 0$ while $f''(x) \geq 0$ a.e. and so by the fundamental theorem of calculus,

$$f'(x) = \varphi'(x) - \varphi'(x_0) = \int_{x_0}^{x} \varphi''(y)\, dy.$$

Hence it follows that $f'(x) \geq 0$ for $x > x_0$ and $f'(x) \leq 0$ for $x < x_0$ and therefore, $f(x) \geq 0$ for all $x \in (a, b)$.

2. Let

$$f(t) := \varphi(u) + t(\varphi(v) - \varphi(u)) - \varphi(u + t(v - u)).$$

Then $f(0) = f(1) = 0$ with $\ddot{f}(t) = -(v - u)^2 \varphi''(u + t(v - u)) \leq 0$ 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. $f(t) \ge 0$. ■

Example 7.32. Taking $\varphi(x) := e^{-x}$, we learn (see Figure 7.3),

$$1 - x \le e^{-x} \text{ for all } x \in \mathbb{R} \tag{7.8}$$

and taking $\varphi(x) = e^{-2x}$ we learn that

$$1 - x \ge e^{-2x} \text{ for } 0 \le x \le 1/2. \tag{7.9}$$

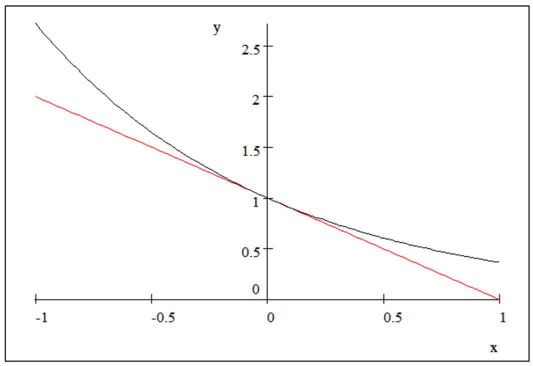


Fig. 7.3. A graph of $1 - x$ and e^{-x} showing that $1 - x \le e^{-x}$ for all x.

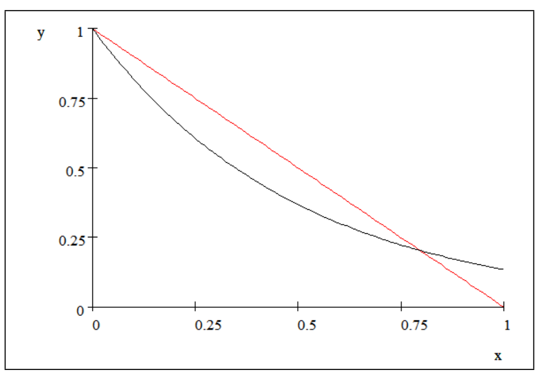


Fig. 7.4. A graph of $1 - x$ and e^{-2x} showing that $1 - x \ge e^{-2x}$ for all $x \in [0, 1/2]$.

Exercise 7.4. 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, $\prod_{n=1}^{N} (1-a_n) \downarrow$ as $N \uparrow$.) Show that if $\{a_n\}_{n=1}^{\infty} \subset [0,1)$, then

$$\prod_{n=1}^{\infty} (1-a_n) = 0 \text{ iff } \sum_{n=1}^{\infty} a_n = \infty.$$

Solution to Exercise (7.4). On one hand we have

$$\prod_{n=1}^{N} (1-a_n) \le \prod_{n=1}^{N} e^{-a_n} = \exp\left(-\sum_{n=1}^{N} a_n\right)$$

which upon passing to the limit as $N \to \infty$ gives

$$\prod_{n=1}^{\infty} (1-a_n) \le \exp\left(-\sum_{n=1}^{\infty} a_n\right).$$

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 $n \ge m$. With this notation we then have for $N \ge m$ that

$$\begin{aligned}\prod_{n=1}^{N} (1-a_n) &= \prod_{n=1}^{m} (1-a_n) \cdot \prod_{n=m+1}^{N} (1-a_n) \\ &\ge \prod_{n=1}^{m} (1-a_n) \cdot \prod_{n=m+1}^{N} e^{-2a_n} = \prod_{n=1}^{m} (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^{N} a_n\right) \\ &\ge \prod_{n=1}^{m} (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^{\infty} a_n\right).\end{aligned}$$

So again letting $N \to \infty$ shows,

$$\prod_{n=1}^{\infty} (1-a_n) \ge \prod_{n=1}^{m} (1-a_n) \cdot \exp\left(-2\sum_{n=m+1}^{\infty} a_n\right) > 0.$$

Lemma 7.33 (Second Borel-Cantelli Lemma). *Suppose that* $\{A_n\}_{n=1}^{\infty}$ *are independent sets. If*

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \tag{7.10}$$

then

$$P(\{A_n \ i.o.\}) = 1. \tag{7.11}$$

Combining this with the first Borel Cantelli Lemma gives the (Borel) Zero-One law,

$$P(A_n \ i.o.) = \begin{cases} 0 \ \textit{if} \sum_{n=1}^{\infty} P(A_n) < \infty \\ 1 \ \textit{if} \sum_{n=1}^{\infty} P(A_n) = \infty \end{cases}.$$

Proof. We are going to prove Eq. (7.11) 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

$$P(\cap_{m \geq k \geq n} A_k^c) = \prod_{m \geq k \geq n} P(A_k^c) = \prod_{m \geq k \geq n} (1 - P(A_k)). \tag{7.12}$$

Using the simple inequality in Eq. (7.8) along with Eq. (7.12) 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. (7.10), we find from the above inequality that $\lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = 0$ and hence

$$P(\cup_{n=1}^{\infty} \cap_{k \geq n} A_k^c) = \lim_{n \to \infty} \lim_{m \to \infty} P(\cap_{m \geq k \geq n} A_k^c) = \lim_{n \to \infty} 0 = 0$$

as desired. ■

Example 7.34 (Example 7.30 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 7.35 (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*

$$\sum_{n=1}^{\infty} P\left(X_1 > \alpha^{-1} c_n\right) = \infty \tag{7.13}$$

while

$$\sum_{n=1}^{\infty} P\left(X_1 > \alpha c_n\right) < \infty. \tag{7.14}$$

Then

$$\limsup_{n\to\infty} \frac{X_n}{c_n} = 1 \;\; a.s. \tag{7.15}$$

Proof. By the second Borel-Cantelli Lemma, Eq. (7.13) 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. (7.14) implies

$$P\left(X_n > \alpha c_n \text{ i.o. } n\right) = 0$$

or equivalently,

$$P\left(X_n \leq \alpha c_n \text{ a.a. } n\right) = 1.$$

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 7.36. Let $\{E_n\}_{n=1}^{\infty}$ be a sequence of independent random variables with exponential distributions determined by

$$P\left(E_n > x\right) = e^{-(x\vee 0)} \text{ or } P\left(E_n \leq x\right) = 1 - e^{-(x\vee 0)}.$$

(Observe that $P\left(E_n \leq 0\right) = 0$) so that $E_n > 0$ a.s.) Then for $c_n > 0$ and $\alpha > 0$, we have

$$\sum_{n=1}^{\infty} P\left(E_n > \alpha c_n\right) = \sum_{n=1}^{\infty} e^{-\alpha c_n} = \sum_{n=1}^{\infty} \left(e^{-c_n}\right)^{\alpha}.$$

Hence if we choose $c_n = \ln n$ so that $e^{-c_n} = 1/n$, then we have

$$\sum_{n=1}^{\infty} P\left(E_n > \alpha \ln n\right) = \sum_{n=1}^{\infty} \left(\frac{1}{n}\right)^{\alpha}$$

which is convergent iff $\alpha > 1$. So by Proposition 7.35, it follows that

$$\limsup_{n\to\infty} \frac{E_n}{\ln n} = 1 \text{ a.s.}$$

Example 7.37. Suppose now that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. distributed by the Poisson distribution with intensity, λ, i.e.

$$P\left(X_1 = k\right) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In this case we have

$$P\left(X_1 \geq n\right) = e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} \geq \frac{\lambda^n}{n!} e^{-\lambda}$$

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

$$\frac{\lambda^n}{n!} e^{-\lambda} \leq P\left(X_1 \geq n\right) \leq \frac{\lambda^n}{n!}.$$

Thus in terms of convergence issues, we may assume that

$$P\left(X_1 \geq x\right) \sim \frac{\lambda^x}{x!} \sim \frac{\lambda^x}{\sqrt{2\pi x} e^{-x} x^x}$$

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

$$P\left(X_1 \geq c_n\right) \sim 1/n.$$

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 that

$$x = \frac{\ln(n)}{\ell_2(n) + O(\ell_3(n))} = \frac{\ln(n)}{\ell_2(n)}\left(1 + O\left(\frac{\ell_3(n)}{\ell_2(n)}\right)\right).$$

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

$$P(X_1 \geq \alpha c_n) \sim \frac{\lambda^{\alpha c_n}}{\sqrt{2\pi\alpha c_n} e^{-\alpha c_n} (\alpha c_n)^{\alpha c_n}} \sim \frac{(\lambda/e)^{\alpha c_n}}{\sqrt{2\pi\alpha c_n}} \frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}}.$$

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,

$$P(X_1 \geq \alpha c_n) \sim \frac{n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}}}{\sqrt{\frac{\ln(n)}{\ell_2(n)}}} \frac{1}{n^{\alpha(1+\varepsilon_n(\alpha))}} = \sqrt{\frac{\ell_2(n)}{\ln(n)}} \frac{1}{n^{\alpha(1+\delta_n(\alpha))}}$$

where $\delta_n(\alpha) \to 0$ as $n \to \infty$. From this observation, we may show,

$$\sum_{n=1}^{\infty} P(X_1 \geq \alpha c_n) < \infty \text{ if } \alpha > 1 \text{ and}$$
$$\sum_{n=1}^{\infty} P(X_1 \geq \alpha c_n) = \infty \text{ if } \alpha < 1$$

and so by Proposition 7.35 we may conclude that

$$\limsup_{n\to\infty} \frac{X_n}{\ln(n)/\ell_2(n)} = 1 \text{ a.s.}$$

7.4 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 7.38. 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 7.39. *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 $A\in\mathcal{F}$.*

Lemma 7.40. *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 $P(X=c)=1$. ■

Proposition 7.41 (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 $A\in\mathcal{T}$.*

Proof. Let $A\in\mathcal{T}\subset\mathcal{B}_\infty$. Since $A\in\mathcal{T}_n$ for all n and $\mathcal{T}_n$ is independent of $\mathcal{B}_n$, it follows that A is independent of $\cup_{n=1}^\infty\mathcal{B}_n$ for all n. Since the latter set is a multiplicative set, it follows that A is independent of $\mathcal{B}_\infty=\sigma(\cup\mathcal{B}_n)=\vee_{n=1}^\infty\mathcal{B}_n$. But $A\in\mathcal{B}$ and hence A is independent of itself, i.e.

$$P(A)=P(A\cap A)=P(A)\,P(A).$$

Since the only $x\in\mathbb{R}$, such that $x=x^2$ is $x=0$ or $x=1$, the result is proved. In particular the tail events in Example 7.38 have probability either 0 or 1. ■

Corollary 7.42. *Keeping the assumptions in Proposition 7.41 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 $c\in\bar{\mathbb{R}}$.*

Let us now suppose that $\Omega:=\mathbb{R}^\infty=\mathbb{R}^{\mathbb{N}}$, $X_n(\omega)=\omega_n$ for all $\omega\in\Omega$, and $\mathcal{B}:=\sigma(X_1,X_2,\dots)$. 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 $T_\pi(\omega)=(\omega_{\pi1},\omega_{\pi2},\dots)$.

Definition 7.43. *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*.*

In the proof below we will use the identities,

$$1_{A \triangle B} = |1_A - 1_B| \text{ and } P(A \triangle B) = \mathbb{E}|1_A - 1_B|.$$

Proposition 7.44 (Hewitt-Savage Zero-One Law). *Let* P *be a probability measure on* $(\Omega, \mathcal{B})$ *such that* $\{X_n\}_{n=1}^\infty$ *is an i.i.d. sequence. Then* $\mathcal{S}$ *is almost trivial.*

Proof. Let $\mathcal{B}_0 := \cup_{n=1}^\infty \sigma(X_1, X_2, \dots, X_n)$. Then $\mathcal{B}_0$ is an algebra and $\sigma(\mathcal{B}_0) = \mathcal{B}$. By the regularity Theorem 5.10, for any $B \in \mathcal{B}$ and $\varepsilon > 0$, there exists $A_n \in \mathcal{B}_0$ such that $A_n \uparrow C \in (\mathcal{B}_0)_\sigma$, $B \subset C$, and $P(C \setminus B) < \varepsilon$. Since

$$\begin{aligned} P(A_n \Delta B) &= P([A_n \setminus B] \cup [B \setminus A_n]) = P(A_n \setminus B) + P(B \setminus A_n) \\ &\to P(C \setminus B) + P(B \setminus C) < \varepsilon, \end{aligned}$$

for sufficiently large n, we have $P(A \Delta B) < \varepsilon$ where $A = A_n \in \mathcal{B}_0$.

Now suppose that $B \in \mathcal{S}$, $\varepsilon > 0$, and $A \in \sigma(X_1, X_2, \dots, X_n) \subset \mathcal{B}_0$ such that $P(A \Delta B) < \varepsilon$. Let $\pi : \mathbb{N} \to \mathbb{N}$ be the permutation defined by $\pi(j) = j + n$, $\pi(j+n) = j$ for $j = 1, 2, \dots, n$, and $\pi(j + 2n) = j + 2n$ for all $j \in \mathbb{N}$. Since

$$B = \{(X_1, \dots, X_n) \in B'\} = \{\omega : (\omega_1, \dots, \omega_n) \in B'\}$$

for some $B' \in \mathcal{B}_{\mathbb{R}^n}$, we have

$$\begin{aligned} T_\pi^{-1}(B) &= \{\omega : ((T_\pi(\omega))_1, \dots, (T_\pi(\omega))_n) \in B'\} \\ &= \{\omega : (\omega_{\pi 1}, \dots, \omega_{\pi n}) \in B'\} \\ &= \{\omega : (\omega_{n+1}, \dots, \omega_{n+n}) \in B'\} \\ &= \{(X_{n+1}, \dots, X_{n+n}) \in B'\} \in \sigma(X_{n+1}, \dots, X_{n+n}), \end{aligned}$$

it follows that B and $T_\pi^{-1}(B)$ are independent with $P(B) = P\left(T_\pi^{-1}(B)\right)$. Therefore $P\left(B \cap T_\pi^{-1} B\right) = P(B)^2$. Combining this observation with the identity, $P(A) = P(A \cap A) = P\left(A \cap T_\pi^{-1} A\right)$, we find

$$\begin{aligned} \left|P(A) - P(B)^2\right| &= \left|P\left(A \cap T_\pi^{-1} A\right) - P\left(B \cap T_\pi^{-1} B\right)\right| = \left|\mathbb{E}\left[1_{A \cap T_\pi^{-1} A} - 1_{B \cap T_\pi^{-1} B}\right]\right| \\ &\le \mathbb{E}\left|1_{A \cap T_\pi^{-1} A} - 1_{B \cap T_\pi^{-1} B}\right| \\ &= \mathbb{E}\left|1_A 1_{T_\pi^{-1} A} - 1_B 1_{T_\pi^{-1} B}\right| \\ &= \mathbb{E}\left|[1_A - 1_B] 1_{T_\pi^{-1} A} + 1_B \left[1_{T_\pi^{-1} A} - 1_{T_\pi^{-1} B}\right]\right| \\ &\le \mathbb{E}|[1_A - 1_B]| + \mathbb{E}\left|1_{T_\pi^{-1} A} - 1_{T_\pi^{-1} B}\right| \\ &= P(A \Delta B) + P\left(T_\pi^{-1} A \Delta T_\pi^{-1} B\right) < 2\varepsilon. \end{aligned}$$

Since $|P(A) - P(B)| \le P(A \Delta B) < \varepsilon$, it follows that

$$\left| P(A) - [P(A) + O(\varepsilon)]^2 \right| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we may conclude that $P(A) = P(A)^2$ for all $A \in \mathcal{S}$. ■

Example 7.45 (Some Random Walk $0-1$ *Law Results).* Continue the notation in Proposition 7.44.

1. As above, if $S_n = X_1 + \cdots + X_n$, 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. The same goes for $\{S_n \in B \text{ a.a.}\}$
2. If $P(X_1 \ne 0) > 0$, then $\limsup_{n\to\infty} S_n = \infty$ a.s. or $\limsup_{n\to\infty} S_n = -\infty$ a.s. Indeed,
$$T_\pi^{-1}\left\{\limsup_{n\to\infty} S_n \le x\right\} = \left\{\limsup_{n\to\infty} S_n \circ T_\pi \le x\right\} = \left\{\limsup_{n\to\infty} S_n \le x\right\}$$
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, a.s.,
$$c = \limsup_{n\to\infty} S_{n+1} = \limsup_{n\to\infty} (S_n + X_1) = \limsup_{n\to\infty} S_n + X_1 = c + X_1,$$
we must have either $c \in \{\pm\infty\}$ 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 $P(X_1 \ne 0) > 0$ and $X_1 \stackrel{\text{d}}{=} -X_1$, i.e. $P(X_1 \in A) = P(-X_1 \in A)$ for all $A \in \mathcal{B}_{\mathbb{R}}$. From item 2. we know that and from what we have already proved, 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 $-X_n \stackrel{\text{d}}{=} X_n$, it follows that $\{X_n\}_{n=1}^\infty \stackrel{\text{d}}{=} \{-X_n\}_{n=1}^\infty$ .The results of Exercise 7.2 then imply that $\limsup_{n\to\infty} S_n \stackrel{\text{d}}{=} \limsup_{n\to\infty} (-S_n)$ and in particular $\limsup_{n\to\infty} (-S_n) = c$ a.s. as well. Thus we have
$$c = \limsup_{n\to\infty} (-S_n) = -\liminf_{n\to\infty} S_n \ge -\limsup_{n\to\infty} S_n = -c.$$
Since the $c = -\infty$ does not satisfy, $c \ge -c$, we must $c = \infty$. Hence in this symmetric case we have shown,

$$\limsup_{n\to\infty} S_n = \infty \text{ and } \limsup_{n\to\infty} (-S_n) = \infty \text{ a.s.}$$

or equivalently that

$$\limsup_{n\to\infty} S_n = \infty \text{ and } \liminf_{n\to\infty} S_n = -\infty \text{ a.s.}$$

8

Integration Theory

In this chapter, we will greatly extend the "simple" integral or expectation which was developed in Section 4.3 above. Recall there that if $(\Omega, \mathcal{B}, \mu)$ was measurable space and $f : \Omega \to [0, \infty]$ was a measurable simple function, then we let

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

8.1 A Quick Introduction to Lebesgue Integration Theory

Theorem 8.1 (Extension to positive functions). *For a positive measurable function, $f : \Omega \to [0, \infty]$, the integral of f with respect to μ is defined by*

$$\int_X f(x) \, d\mu(x) := \sup \{\mathbb{E}_\mu \varphi : \varphi \text{ is simple and } \varphi \le f\}.$$

This integral has the following properties.

1. *This integral is linear in the sense that*
$$\int_\Omega (f + \lambda g) \, d\mu = \int_\Omega f d\mu + \lambda \int_\Omega g d\mu$$
whenever $f, g \ge 0$ are measurable functions and $\lambda \in [0, \infty)$.
2. *The integral is continuous under increasing limits, i.e. if $0 \le f_n \uparrow f$, then*
$$\int_\Omega f \ d\mu = \int_\Omega \lim_{n\to\infty} f_n \ d\mu = \lim_{n\to\infty} \int_\Omega f_n \ d\mu.$$
See the monotone convergence Theorem 8.15 below.

Remark 8.2. Given $f : \Omega \to [0, \infty]$ measurable, we know from the approximation Theorem 6.32 $\varphi_n \uparrow f$ where

$$\varphi_n := \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} 1_{\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\}} + n 1_{\{f > n2^n\}}.$$

Therefore by the monotone convergence theorem,

$$\begin{aligned}\int_\Omega f d\mu &= \lim_{n\to\infty}\int_\Omega \varphi_n d\mu \\ &= \lim_{n\to\infty}\left[\sum_{k=0}^{n2^n-1}\frac{k}{2^n}\mu\left(\frac{k}{2^n}<f\le\frac{k+1}{2^n}\right)+n\mu\left(f>n2^n\right)\right].\end{aligned}$$

We call a function, $f:\Omega\to\bar{\mathbb{R}}$, **integrable** if it is measurable and $\int_\Omega |f|\,d\mu<\infty$. We will denote the space of μ –integrable functions by $L^1(\mu)$

Theorem 8.3 (Extension to integrable functions). *The integral extends to a linear function from $L^1(\mu)\to\mathbb{R}$. Moreover this extension is continuous under dominated convergence (see Theorem 8.34). That is if $f_n\in L^1(\mu)$ and there exists $g\in L^1(\mu)$ such that $|f_n|\le g$ and $f:=\lim_{n\to\infty}f_n$ exists pointwise, then*

$$\int_\Omega f\ d\mu=\int_\Omega \lim_{n\to\infty}f_n\ d\mu=\lim_{n\to\infty}f_n\int_\Omega\ d\mu.$$

Notation 8.4 *We write $\int_A f d\mu := \int_\Omega 1_A f\ d\mu$ for all $A\in\mathcal{B}$ where f is a measurable function such that $1_A f$ is either non-negative or integrable.*

Notation 8.5 *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 8.6. Suppose $-\infty<a<b<\infty$, $f\in C([a,b],\mathbb{R})$ and m be Lebesgue measure on $\mathbb{R}$. Given a partition,

$$\pi=\{a=a_0<a_1<\cdots<a_n=b\},$$

let

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

and

$$f_\pi(x):=\sum_{l=0}^{n-1}f(a_l)\,1_{(a_l,a_{l+1}]}(x).$$

Then

$$\int_a^b f_\pi\ dm=\sum_{l=0}^{n-1}f(a_l)\,m\left((a_l,a_{l+1}]\right)=\sum_{l=0}^{n-1}f(a_l)\left(a_{l+1}-a_l\right)$$

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty}\text{mesh}(\pi_k)=0$, we know that

$$\lim_{k\to\infty}\int_a^b f_{\pi_k}\ dm=\int_a^b f(x)\,dx \tag{8.1}$$

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 $|f_{\pi_k}(x)| \leq g(x) := M1_{(a,b]}(x)$ 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. (8.1) 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 8.51 below for a more general statement along these lines.

Theorem 8.7 (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 $\left|1_{(a,x)}(y)f(y)\right| \leq 1_{(a,b)}(y)\,|f(y)|$ is an L^1 – function, it follows from the dominated convergence Theorem 8.34 that F is continuous on $[a,b]$. Simple manipulations show,

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

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

$$\int_a^b f(x)dm(x) = F(b) = F(b) - F(a)$$
$$= (G(b) + C) - (G(a) + C) = G(b) - G(a).$$

■

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

Example 8.8. 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 8.6 and the fundamental theorem of calculus for Riemann integrals (or Theorem 8.7 below),

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

and

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

Let us also consider the functions x^{-p},

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

If $p = 1$ we find

$$\int_{(0,1]} \frac{1}{x^p}\, dm(x) = \lim_{n\to\infty} \int_{\frac{1}{n}}^1 \frac{1}{x} dx = \lim_{n\to\infty} \ln(x)|_{1/n}^1 = \infty.$$

Exercise 8.1. 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 8.9. The following limit holds,

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

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 $x\geq 0$ and by taking logarithms of Eq. (7.8),

$$\ln(1-x)\leq -x \text{ for } x<1.$$

Therefore, for $x<n$, we have

$$\left(1-\frac{x}{n}\right)^n=e^{n\ln\left(1-\frac{x}{n}\right)}\leq e^{-n\left(\frac{x}{n}\right)}=e^{-x}$$

from which it follows that

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

From Example 8.8, 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,

$$\begin{aligned}\lim_{n\to\infty}\int_0^n\left(1-\frac{x}{n}\right)^n dm(x)&=\lim_{n\to\infty}\int_0^\infty f_n(x)dm(x)\\&=\int_0^\infty\lim_{n\to\infty}f_n(x)dm(x)=\int_0^\infty e^{-x}dm(x)=1.\end{aligned}$$

The limit in the above example 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 $n>x$,

$$\begin{aligned}\frac{d}{dn}\ln f_n(x)&=\frac{d}{dn}\left[n\ln\left(1-\frac{x}{n}\right)\right]=\ln\left(1-\frac{x}{n}\right)+\frac{n}{1-\frac{x}{n}}\frac{x}{n^2}\\&=\ln\left(1-\frac{x}{n}\right)+\frac{\frac{x}{n}}{1-\frac{x}{n}}=h(x/n)\end{aligned}$$

where, for $0\leq y<1$,

$$h(y):=\ln(1-y)+\frac{y}{1-y}.$$

Since $h(0)=0$ and

$$h'(y)=-\frac{1}{1-y}+\frac{1}{1-y}+\frac{y}{(1-y)^2}>0$$

it follows that $h\geq 0$. Thus we have shown, $f_n(x)\uparrow e^{-x}$ as $n\to\infty$ as claimed.

Example 8.10 (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.,

$$\lim_{n\to\infty}\int_0^{\pi}\cos\left(\sin\frac{\theta}{n}\right)e^{-n\sin(\theta)}d\theta = \int_{\mathbb{R}} 1_{\{\pi\}}(\theta)\,dm(\theta) = m(\{\pi\}) = 0.$$

Exercise 8.2 (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,

$$f(a) := \lim_{n\to\infty}\int_a^{\infty} n(1+n^2x^2)^{-1}dx.$$

Now that we have an overview of the Lebesgue integral, let us proceed to the formal development of the facts stated above.

8.2 Integrals of positive functions

Definition 8.11. *Let $L^+ = L^+(\mathcal{B}) = \{f : X \to [0,\infty] : f \text{ is measurable}\}$. Define*

$$\int_X f(x)\,d\mu(x) = \int_X f d\mu := \sup\{\mathbb{E}_\mu\varphi : \varphi \text{ is simple and } \varphi \le f\}.$$

*We say the $f\in L^+$ is **integrable** if $\int_X f d\mu < \infty$. If $A\in\mathcal{B}$, let*

$$\int_A f(x)\,d\mu(x) = \int_A f d\mu := \int_X 1_A f\ d\mu.$$

Remark 8.12. Because of item 3. of Proposition 4.16, if φ is a non-negative simple function, $\int_X \varphi d\mu = \mathbb{E}_\mu \varphi$ so that $\int_X$ is an extension of $\mathbb{E}_\mu$.

Lemma 8.13. *Let $f, g \in L^+(\mathcal{B})$. Then:*

1. *if $\lambda \geq 0$, then*
$$\int_X \lambda f d\mu = \lambda \int_X f d\mu$$
wherein $\lambda \int_X f d\mu \equiv 0$ if $\lambda = 0$, even if $\int_X f d\mu = \infty$.
2. *if $0 \leq f \leq g$, then*
$$\int_X f d\mu \leq \int_X g d\mu. \tag{8.2}$$
3. *For all $\varepsilon > 0$ and $p > 0$,*
$$\mu(f \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int_X f^p 1_{\{f \geq \varepsilon\}} d\mu \leq \frac{1}{\varepsilon^p} \int_X f^p d\mu. \tag{8.3}$$
The inequality in Eq. (8.3) is called Chebyshev's Inequality for $p = 1$ and Markov's inequality for $p = 2$.
4. *If $\int_X f d\mu < \infty$ 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,

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

2. Since

$$\{\varphi \text{ is simple and } \varphi \leq f\} \subset \{\varphi \text{ is simple and } \varphi \leq g\},$$

Eq. (8.2) follows from the definition of the integral.

3. Since $1_{\{f \geq \varepsilon\}} \leq 1_{\{f \geq \varepsilon\}} \frac{1}{\varepsilon} f \leq \frac{1}{\varepsilon} f$ we have

$$1_{\{f \geq \varepsilon\}} \leq 1_{\{f \geq \varepsilon\}} \left(\frac{1}{\varepsilon} f\right)^p \leq \left(\frac{1}{\varepsilon} f\right)^p$$

and by monotonicity and the multiplicative property of the integral,

$$\mu(f \geq \varepsilon) = \int_X 1_{\{f \geq \varepsilon\}} d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_X 1_{\{f \geq \varepsilon\}} f^p d\mu \leq \left(\frac{1}{\varepsilon}\right)^p \int_X f^p d\mu.$$

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

$$n\mu(f=\infty) = \mathbb{E}_\mu(\varphi_n) \le \int_X f d\mu$$

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

Moreover,

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

with $\mu(f>1/n) \le n\int_X f d\mu < \infty$ for each n. ■

Lemma 8.14 (Sums as Integrals). *Let X be a set and $\rho : X \to [0,\infty]$ be a function, let $\mu = \sum_{x\in X}\rho(x)\delta_x$ on $\mathcal{B} = 2^X$, i.e.*

$$\mu(A) = \sum_{x\in A}\rho(x).$$

If $f : X\to[0,\infty]$ is a function (which is necessarily measurable), then

$$\int_X f d\mu = \sum_X f\rho.$$

Proof. Suppose that $\varphi : X\to[0,\infty)$ is a simple function, then $\varphi = \sum_{z\in[0,\infty)} z1_{\{\varphi=z\}}$ and

$$\begin{aligned}\sum_X \varphi\rho &= \sum_{x\in X}\rho(x)\sum_{z\in[0,\infty)} z1_{\{\varphi=z\}}(x) = \sum_{z\in[0,\infty)} z\sum_{x\in X}\rho(x)1_{\{\varphi=z\}}(x)\\ &= \sum_{z\in[0,\infty)} z\mu(\{\varphi=z\}) = \int_X \varphi d\mu.\end{aligned}$$

So if $\varphi : X\to[0,\infty)$ is a simple function such that $\varphi\le f$, then

$$\int_X \varphi d\mu = \sum_X \varphi\rho \le \sum_X f\rho.$$

Taking the sup over φ in this last equation then shows that

$$\int_X f d\mu \le \sum_X f\rho.$$

For the reverse inequality, let $\Lambda \subset\subset X$ be a finite set and $N\in(0,\infty)$. Set $f^N(x) = \min\{N, f(x)\}$ and let $\varphi_{N,\Lambda}$ be the simple function given by $\varphi_{N,\Lambda}(x) := 1_\Lambda(x) f^N(x)$. Because $\varphi_{N,\Lambda}(x)\le f(x)$,

$$\sum_\Lambda f^N\rho = \sum_X \varphi_{N,\Lambda}\rho = \int_X \varphi_{N,\Lambda} d\mu \le \int_X f d\mu.$$

Since $f^N \uparrow f$ as $N \to \infty$, we may let $N \to \infty$ in this last equation to concluded

$$\sum_{\Lambda} f\rho \leq \int_X f d\mu.$$

Since Λ is arbitrary, this implies

$$\sum_{X} f\rho \leq \int_X f d\mu.$$

■

Theorem 8.15 (Monotone Convergence Theorem). *Suppose $f_n \in L^+$ is a sequence of functions such that $f_n \uparrow f$ (f is necessarily in L^+) then*

$$\int f_n \uparrow \int f \text{ as } n \to \infty.$$

Proof. Since $f_n \leq f_m \leq f$, for all $n \leq m < \infty$,

$$\int f_n \leq \int f_m \leq \int f$$

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

$$\lim_{n\to\infty} \int f_n \leq \int f. \tag{8.4}$$

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

$$\int f_n \geq \mathbb{E}_\mu [\alpha 1_{X_n}\varphi] = \alpha \mathbb{E}_\mu [1_{X_n}\varphi]. \tag{8.5}$$

Then using the continuity of μ under increasing unions,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}_\mu [1_{X_n}\varphi] &= \lim_{n\to\infty} \int 1_{X_n} \sum_{y>0} y 1_{\{\varphi=y\}} \\
&= \lim_{n\to\infty} \sum_{y>0} y\mu(X_n \cap \{\varphi = y\}) \\
&\overset{\text{finite sum}}{=} \sum_{y>0} y \lim_{n\to\infty} \mu(X_n \cap \{\varphi = y\}) \\
&= \sum_{y>0} y \lim_{n\to\infty} \mu(\{\varphi = y\}) = \mathbb{E}_\mu [\varphi]
\end{aligned}$$

This identity allows us to let $n \to \infty$ in Eq. (8.5) to conclude $\lim_{n\to\infty} \int f_n \geq \alpha \mathbb{E}_\mu [\varphi]$ and since $\alpha \in (0,1)$ was arbitrary we may further conclude,$\mathbb{E}_\mu [\varphi] \leq$

$\lim_{n\to\infty}\int f_n$. The latter inequality being true for all simple functions φ with $\varphi \le f$ then implies that

$$\int f \le \lim_{n\to\infty}\int f_n,$$

which combined with Eq. (8.4) proves the theorem. ■

Corollary 8.16. *If $f_n \in L^+$ is a sequence of functions then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty}\int f_n.$$

In particular, if $\sum_{n=1}^{\infty}\int f_n < \infty$ then $\sum_{n=1}^{\infty} f_n < \infty$ a.e.

Proof. First off we show that

$$\int (f_1 + f_2) = \int f_1 + \int f_2$$

by choosing non-negative simple function φ_n and ψ_n such that $\varphi_n \uparrow f_1$ and $\psi_n \uparrow f_2$. Then $(\varphi_n + \psi_n)$ is simple as well and $(\varphi_n + \psi_n) \uparrow (f_1 + f_2)$ so by the monotone convergence theorem,

$$\begin{aligned}\int (f_1 + f_2) &= \lim_{n\to\infty}\int (\varphi_n + \psi_n) = \lim_{n\to\infty}\left(\int \varphi_n + \int \psi_n\right) \\ &= \lim_{n\to\infty}\int \varphi_n + \lim_{n\to\infty}\int \psi_n = \int f_1 + \int f_2.\end{aligned}$$

Now to the general case. Let $g_N := \sum\limits_{n=1}^{N} f_n$ and $g = \sum\limits_{1}^{\infty} f_n$, then $g_N \uparrow g$ and so again by monotone convergence theorem and the additivity just proved,

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

■

Remark 8.17. It is in the proof of this corollary (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 $f : X \to [0,\infty]$ 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 8.16, we use the approximation Theorem 6.32 which relies heavily on the measurability of the functions to be approximated.

Example 8.18. Suppose, $\Omega = \mathbb{N}$, $\mathcal{B} := 2^{\mathbb{N}}$, and $\mu(A) = \#(A)$ for $A \subset \Omega$ is the counting measure on $\mathcal{B}$. Then for $f : \mathbb{N} \to [0, \infty)$, the function

$$f_N(\cdot) := \sum_{n=1}^{N} f(n)\, 1_{\{n\}}$$

is a simple function with $f_N \uparrow f$ as $N \to \infty$. So by the monotone convergence theorem,

$$\begin{aligned}\int_{\mathbb{N}} f d\mu &= \lim_{N\to\infty} \int_{\mathbb{N}} f_N d\mu = \lim_{N\to\infty} \sum_{n=1}^{N} f(n)\, \mu(\{n\}) \\ &= \lim_{N\to\infty} \sum_{n=1}^{N} f(n) = \sum_{n=1}^{\infty} f(n).\end{aligned}$$

Exercise 8.3. 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. **Hint:** use Example 8.18 and the monotone convergence theorem.

Proposition 8.19. *Suppose that $f \geq 0$ is a measurable function. Then $\int_X f d\mu = 0$ iff $f = 0$ a.e. Also if $f, g \geq 0$ are measurable functions such that $f \leq g$ a.e. then $\int f d\mu \leq \int g d\mu$. In particular if $f = g$ a.e. then $\int f d\mu = \int g d\mu$.*

Proof. If $f = 0$ a.e. and $\varphi \leq f$ is a simple function then $\varphi = 0$ a.e. This implies that $\mu(\varphi^{-1}(\{y\})) = 0$ for all $y > 0$ and hence $\int_X \varphi d\mu = 0$ and therefore $\int_X f d\mu = 0$. Conversely, if $\int f d\mu = 0$, then by (Lemma 8.13),

$$\mu(f \geq 1/n) \leq n \int f d\mu = 0 \text{ for all } n.$$

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. $E := \{x \in X : f(x) > g(x)\}$. 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 8.20. *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 $E \subset X$ 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 8.19,

$$\int f_n = \int f_n 1_{E^c} \uparrow \int f 1_{E^c} = \int f \text{ as } n \to \infty.$$

■

Lemma 8.21 (Fatou's Lemma). *If $f_n : X \to [0, \infty]$ is a sequence of measurable functions then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n$$

Proof. Define $g_k := \inf_{n \ge k} f_n$ so that $g_k \uparrow \liminf_{n\to\infty} f_n$ as $k \to \infty$. Since $g_k \le f_n$ for all $k \le n$,

$$\int g_k \le \int f_n \text{ for all } n \ge k$$

and therefore

$$\int g_k \le \lim \inf_{n\to\infty} \int f_n \text{ for all } k.$$

We may now use the monotone convergence theorem to let $k \to \infty$ to find

$$\int \lim \inf_{n\to\infty} f_n = \int \lim_{k\to\infty} g_k \overset{\text{MCT}}{=} \lim_{k\to\infty} \int g_k \le \lim \inf_{n\to\infty} \int f_n.$$

■

The following Lemma and the next Corollary are simple applications of Corollary 8.16.

Lemma 8.22 (The First Borell – Carntelli Lemma). *Let $(X, \mathcal{B}, \mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

$$\{A_n \text{ i.o.}\} = \{x \in X : x \in A_n \text{ for infinitely many } n\text{'s}\} = \bigcap_{N=1}^{\infty} \bigcup_{n \ge N} A_n.$$

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

$$\{A_n \text{ i.o.}\} = \left\{ x \in X : \sum_{n=1}^{\infty} 1_{A_n}(x) = \infty \right\}.$$

Hence if $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then

$$\infty > \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \int_X 1_{A_n} \, d\mu = \int_X \sum_{n=1}^{\infty} 1_{A_n} \, d\mu$$

implies that $\sum_{n=1}^{\infty} 1_{A_n}(x) < \infty$ for μ - a.e. x. That is to say $\mu(\{A_n \text{ i.o.}\}) = 0$. (Second Proof.) Of course we may give a strictly measure theoretic proof of this fact:

$$\begin{aligned}\mu(A_n \text{ i.o.}) &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} A_n\right) \\ &\leq \lim_{N\to\infty} \sum_{n\geq N} \mu(A_n)\end{aligned}$$

and the last limit is zero since $\sum_{n=1}^{\infty} \mu(A_n) < \infty$. ■

Corollary 8.23. *Suppose that $(X, \mathcal{B}, \mu)$ 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

$$\begin{aligned}\mu\left(\cup_{n=1}^{\infty} A_n\right) &= \int_X 1_{\cup_{n=1}^{\infty} A_n} d\mu \text{ and} \\ \sum_{n=1}^{\infty} \mu(A_n) &= \int_X \sum_{n=1}^{\infty} 1_{A_n} d\mu\end{aligned}$$

it suffices to show

$$\sum_{n=1}^{\infty} 1_{A_n} = 1_{\cup_{n=1}^{\infty} A_n} \ \mu - \text{a.e.} \tag{8.6}$$

Now $\sum_{n=1}^{\infty} 1_{A_n} \geq 1_{\cup_{n=1}^{\infty} A_n}$ and $\sum_{n=1}^{\infty} 1_{A_n}(x) \neq 1_{\cup_{n=1}^{\infty} A_n}(x)$ iff $x \in A_i \cap A_j$ for some $i \neq j$, that is

$$\left\{x : \sum_{n=1}^{\infty} 1_{A_n}(x) \neq 1_{\cup_{n=1}^{\infty} A_n}(x)\right\} = \cup_{i<j} A_i \cap A_j$$

and the latter set has measure 0 being the countable union of sets of measure zero. This proves Eq. (8.6) and hence the corollary. ■

Example 8.24. 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 8.7,

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

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 $[0,1]$.

8.3 Integrals of Complex Valued Functions

Definition 8.25. *A measurable function $f : X \to \bar{\mathbb{R}}$ is* ***integrable*** *if $f_+ := f 1_{\{f\ge 0\}}$ and $f_- = -f\, 1_{\{f\le 0\}}$ 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*

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu$$

Convention: If $f, g : X \to \bar{\mathbb{R}}$ are two measurable functions, let $f+g$ denote the collection of measurable functions $h : X \to \bar{\mathbb{R}}$ such that $h(x) = f(x)+g(x)$ whenever $f(x)+g(x)$ 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 8.26 (Abuse of notation) *We will sometimes denote the integral $\int_X f d\mu$ by $\mu(f)$. With this notation we have $\mu(A) = \mu(1_A)$ for all $A \in \mathcal{B}$.*

Remark 8.27. Since

$$f_\pm \le |f| \le f_+ + f_-,$$

a measurable function f is **integrable** iff $\int |f|\ d\mu < \infty$. Hence

$$\mathrm{L}^1(\mu;\mathbb{R}) := \left\{ f : X \to \bar{\mathbb{R}} :\ f \text{ is measurable and } \int_X |f|\ d\mu < \infty \right\}.$$

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 8.19 that $\int f d\mu = \int g d\mu$. In particular if $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ we may define

$$\int_X (f+g)\, d\mu = \int_X h d\mu$$

where h is any element of $f+g$.

Proposition 8.28. *The map*

$$f \in \mathrm{L}^1(\mu;\mathbb{R}) \to \int_X f d\mu \in \mathbb{R}$$

is linear and has the monotonicity property: $\int f d\mu \le \int g d\mu$ *for all* $f, g \in \mathrm{L}^1(\mu;\mathbb{R})$ *such that* $f \le g$ *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

$$|af + bg| \le |a|\,|f| + |b|\,|g| \in \mathrm{L}^1(\mu;\mathbb{R}).$$

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 8.19. ■

Definition 8.29. *A measurable function $f : X \to \mathbb{C}$ is **integrable** if $\int_X |f| \; d\mu < \infty$. Analogously to the real case, let*

$$\mathrm{L}^1(\mu;\mathbb{C}) := \left\{ f : X \to \mathbb{C} : \; f \text{ is measurable and } \int_X |f| \; d\mu < \infty \right\}.$$

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 $f : X \to [0,\infty]$ is a measurable function, let

$$\int_A f d\mu := \int_X 1_A f d\mu.$$

Proposition 8.30. *Suppose that $f \in \mathrm{L}^1(\mu;\mathbb{C})$, then*

$$\left| \int_X f d\mu \right| \le \int_X |f| \, d\mu. \tag{8.7}$$

Proof. Start by writing $\int_X f \; d\mu = Re^{i\theta}$ with $R \ge 0$. We may assume that $R = \left|\int_X f d\mu\right| > 0$ since otherwise there is nothing to prove. Since

$$R = e^{-i\theta} \int_X f \; d\mu = \int_X e^{-i\theta} f \; d\mu = \int_X \mathrm{Re}\left(e^{-i\theta} f\right) d\mu + i \int_X \mathrm{Im}\left(e^{-i\theta} f\right) d\mu,$$

it must be that $\int_X \mathrm{Im}\left[e^{-i\theta} f\right] d\mu = 0$. Using the monotonicity in Proposition 8.19,

$$\left| \int_X f d\mu \right| = \int_X \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_X \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_X |f| \, d\mu.$$

■

Proposition 8.31. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \neq 0\}$ is σ – finite, in fact $\{|f| \geq \frac{1}{n}\} \uparrow \{f \neq 0\}$ and $\mu(|f| \geq \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) $\int_E f = \int_E g$ for all $E \in \mathcal{B}$
   b) $\int\limits_X |f - g| = 0$
   c) $f = g$ a.e.

Proof. 1. By Chebyshev's inequality, Lemma 8.13,

$$\mu(|f| \geq \frac{1}{n}) \leq n \int_X |f| \, d\mu < \infty$$

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 $1_E \,\mathrm{Re}(f - g) \geq 0$, 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

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

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) $\Longrightarrow$ (b) is clear and so is (b) $\Longrightarrow$ (a) since

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

■

Definition 8.32. *Let $(X, \mathcal{B}, \mu)$ be a measure space and $L^1(\mu) = L^1(X, \mathcal{B}, \mu)$ 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 8.33. More generally we may define $L^p(\mu) = L^p(X, \mathcal{B}, \mu)$ for $p \in [1, \infty)$ as the set of measurable functions f such that

$$\int_X |f|^p \, d\mu < \infty$$

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.

Theorem 8.34 (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* $\int_X g_n d\mu \to \int_X g d\mu$. *Then* $f \in \mathrm{L}^1(\mu)$ *and*

$$\int_X f d\mu = \lim_{h \to \infty} \int_X f_n d\mu.$$

(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,

$$\begin{aligned}
\int_X (g \pm f) d\mu &= \int_X \liminf_{n\to\infty} (g_n \pm f_n)\, d\mu \le \liminf_{n\to\infty} \int_X (g_n \pm f_n)\, d\mu \\
&= \lim_{n\to\infty} \int_X g_n d\mu + \liminf_{n\to\infty} \left(\pm \int_X f_n d\mu\right) \\
&= \int_X g d\mu + \liminf_{n\to\infty} \left(\pm \int_X f_n d\mu\right)
\end{aligned}$$

Since $\liminf_{n\to\infty}(-a_n) = -\limsup\limits_{n\to\infty} a_n$, we have shown,

$$\int_X g d\mu \pm \int_X f d\mu \le \int_X g d\mu + \begin{cases} \liminf_{n\to\infty} \int_X f_n d\mu \\ -\limsup\limits_{n\to\infty} \int_X f_n d\mu \end{cases}$$

and therefore

$$\limsup_{n\to\infty} \int_X f_n d\mu \le \int_X f d\mu \le \liminf_{n\to\infty} \int_X f_n d\mu.$$

This shows that $\lim\limits_{n\to\infty} \int_X f_n d\mu$ exists and is equal to $\int_X f d\mu$. ∎

Exercise 8.4. Give another proof of Proposition 8.30 by first proving Eq. (8.7) with f being a simple function in which case the triangle inequality for complex numbers will do the trick. Then use the approximation Theorem 6.32 along with the dominated convergence Theorem 8.34 to handle the general case.

Proposition 8.35. *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. By definition, $\{Z_j\}_{j=1}^n$ are independent iff $\{\sigma(Z_j)\}_{j=1}^n$ are independent. Then as we have seen in a homework problem,

$$\mathbb{E}[1_{A_1}\dots 1_{A_n}] = \mathbb{E}[1_{A_1}]\dots\mathbb{E}[1_{A_n}] \text{ when } A_i \in \sigma(Z_i) \text{ for each } i.$$

By multi-linearity it follows that

$$\mathbb{E}[\varphi_1\dots\varphi_n] = \mathbb{E}[\varphi_1]\dots\mathbb{E}[\varphi_n]$$

whenever φ_i are bounded $\sigma(Z_i)$ – measurable simple functions. By approximation by simple functions and the monotone and dominated convergence theorem,

$$\mathbb{E}[Y_1\dots Y_n] = \mathbb{E}[Y_1]\dots\mathbb{E}[Y_n]$$

whenever Y_i is $\sigma(Z_i)$ – measurable and either $Y_i \geq 0$ or Y_i is bounded. Taking $Y_i = |Z_i|$ then implies that

$$\mathbb{E}\left[\prod_{j=1}^n |Z_j|\right] = \prod_{j=1}^n \mathbb{E}|Z_j| < \infty$$

so that $\prod_{j=1}^n Z_j$ is integrable. Moreover, for $K > 0$, let $Z_i^K = Z_i 1_{|Z_i|\leq K}$, then

$$\mathbb{E}\left[\prod_{j=1}^n Z_j 1_{|Z_j|\leq K}\right] = \prod_{j=1}^n \mathbb{E}\left[Z_j 1_{|Z_j|\leq K}\right].$$

Now apply the dominated convergence theorem, $n+1$ – times, to conclude

$$\mathbb{E}\left[\prod_{j=1}^n Z_j\right] = \lim_{K\to\infty}\mathbb{E}\left[\prod_{j=1}^n Z_j 1_{|Z_j|\leq K}\right] = \prod_{j=1}^n \lim_{K\to\infty}\mathbb{E}\left[Z_j 1_{|Z_j|\leq K}\right] = \prod_{j=1}^n \mathbb{E}Z_j.$$

The dominating functions used here are $\prod_{j=1}^n |Z_j|$, and $\{|Z_j|\}_{j=1}^n$ respectively.
■

Corollary 8.36. *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*

$$\int_X \left(\sum_{n=1}^\infty f_n\right) d\mu = \sum_{n=1}^\infty \int_X f_n d\mu.$$

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,

$$\begin{aligned}\int_X \left(\sum_{n=1}^\infty f_n\right) d\mu &= \int_X \lim_{N\to\infty} S_N d\mu = \lim_{N\to\infty} \int_X S_N d\mu \\ &= \lim_{N\to\infty} \sum_{n=1}^N \int_X f_n d\mu = \sum_{n=1}^\infty \int_X f_n d\mu.\end{aligned}$$

■

Example 8.37 (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 8.36 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 $r = \max(|\beta|, |\alpha|)$.

Corollary 8.38 (Differentiation Under the Integral). *Suppose that* $J \subset \mathbb{R}$ *is an open interval and* $f : J \times X \to \mathbb{C}$ *is a function such that*

1. $x \to f(t, x)$ *is measurable for each* $t \in J$.
2. $f(t_0, \cdot) \in L^1(\mu)$ *for some* $t_0 \in J$.
3. $\frac{\partial f}{\partial t}(t, x)$ *exists for all* (t, x).

4. There is a function $g \in L^1(\mu)$ such that $\left|\frac{\partial f}{\partial t}(t,\cdot)\right| \le g$ for each $t \in J$.

Then $f(t,\cdot) \in L^1(\mu)$ for all $t \in J$ (i.e. $\int_X |f(t,x)|\, d\mu(x) < \infty$), $t \to \int_X f(t,x)d\mu(x)$ is a differentiable function on J and

$$\frac{d}{dt}\int_X f(t,x)d\mu(x) = \int_X \frac{\partial f}{\partial t}(t,x)d\mu(x).$$

Proof. By considering the real and imaginary parts of f separately, we may assume that f is real. Also notice that

$$\frac{\partial f}{\partial t}(t,x) = \lim_{n\to\infty} n(f(t+n^{-1},x) - f(t,x))$$

and therefore, for $x \to \frac{\partial f}{\partial t}(t,x)$ is a sequential limit of measurable functions and hence is measurable for all $t \in J$. By the mean value theorem,

$$|f(t,x) - f(t_0,x)| \le g(x)\,|t - t_0| \text{ for all } t \in J \tag{8.8}$$

and hence

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

This shows $f(t,\cdot) \in L^1(\mu)$ for all $t \in J$. Let $G(t) := \int_X f(t,x)d\mu(x)$, then

$$\frac{G(t) - G(t_0)}{t - t_0} = \int_X \frac{f(t,x) - f(t_0,x)}{t - t_0} d\mu(x).$$

By assumption,

$$\lim_{t\to t_0} \frac{f(t,x) - f(t_0,x)}{t - t_0} = \frac{\partial f}{\partial t}(t,x) \text{ for all } x \in X$$

and by Eq. (8.8),

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

Therefore, we may apply the dominated convergence theorem to conclude

$$\begin{aligned}\lim_{n\to\infty} \frac{G(t_n) - G(t_0)}{t_n - t_0} &= \lim_{n\to\infty}\int_X \frac{f(t_n,x) - f(t_0,x)}{t_n - t_0} d\mu(x)\\ &= \int_X \lim_{n\to\infty} \frac{f(t_n,x) - f(t_0,x)}{t_n - t_0} d\mu(x)\\ &= \int_X \frac{\partial f}{\partial t}(t_0,x)d\mu(x)\end{aligned}$$

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

$$\dot{G}(t_0) = \int_X \frac{\partial f}{\partial t}(t_0,x)d\mu(x).$$

■

Example 8.39. Recall from Example 8.8 that

$$\lambda^{-1} = \int_{[0,\infty)} e^{-\lambda x} dm(x) \text{ for all } \lambda > 0.$$

Let $\varepsilon > 0$. For $\lambda \geq 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

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

Using this fact, Corollary 8.38 and induction gives

$$\begin{aligned} n!\lambda^{-n-1} &= \left(-\frac{d}{d\lambda}\right)^n \lambda^{-1} = \int_{[0,\infty)} \left(-\frac{d}{d\lambda}\right)^n e^{-\lambda x} dm(x) \\ &= \int_{[0,\infty)} x^n e^{-\lambda x} dm(x). \end{aligned}$$

That is $n! = \lambda^n \int_{[0,\infty)} x^n e^{-\lambda x} dm(x)$. Recall that

$$\Gamma(t) := \int_{[0,\infty)} x^{t-1} e^{-x} dx \text{ for } t > 0.$$

(The reader should check that $\Gamma(t) < \infty$ for all $t > 0$.) We have just shown that $\Gamma(n+1) = n!$ for all $n \in \mathbb{N}$.

Remark 8.40. Corollary 8.38 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 8.38 to $g(t) := \int_0^\infty 1_{x \leq t} dm(x)$,

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

The last integral is zero since $\frac{\partial}{\partial t} 1_{x \leq t} = 0$ 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 8.38 has been violated in this example.)

8.4 Densities and Change of Variables Theorems

Exercise 8.5. 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

$$\int_X f d\nu = \int_X f \rho d\mu. \tag{8.9}$$

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. (8.9) still holds.

Solution to Exercise (8.5). The fact that ν is a measure follows easily from Corollary 8.16. Clearly Eq. (8.9) 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,

$$\begin{aligned} \int_X f d\nu &= \int_X f_+ d\nu - \int_X f_- d\nu = \int_X f_+ \rho d\mu - \int_X f_- \rho d\mu \\ &= \int_X \left[f_+\rho - f_-\rho\right] d\mu = \int_X f \rho d\mu. \end{aligned}$$

The complex case easily follows from this identity.

Notation 8.41 *It is customary to informally describe ν defined in Exercise 8.5 by writing $d\nu = \rho d\mu$.*

Exercise 8.6. 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 $A \in \mathcal{F}$.

1. Show ν is a measure. (We will write $\nu = f_*\mu$ or $\nu = \mu \circ f^{-1}$.)
2. Show
$$\int_Y g d\nu = \int_X (g \circ f)\, d\mu \tag{8.10}$$
for all measurable functions $g : Y \to [0, \infty]$. **Hint:** see the hint from Exercise 8.5.
3. 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. (8.10) holds for all $g \in \mathrm{L}^1(\nu)$.

Solution to Exercise (8.6). The fact that ν is a measure is a direct check which will be left to the reader. The key computation is to observe that if $A \in \mathcal{F}$ and $g = 1_A$, then

$$\int_Y g d\nu = \int_Y 1_A d\nu = \nu(A) = \mu\left(f^{-1}(A)\right) = \int_X 1_{f^{-1}(A)} d\mu.$$

Moreover, $1_{f^{-1}(A)}(x) = 1$ iff $x \in f^{-1}(A)$ which happens iff $f(x) \in A$ and hence $1_{f^{-1}(A)}(x) = 1_A(f(x)) = g(f(x))$ for all $x \in X$. Therefore we have

$$\int_Y g d\nu = \int_X (g \circ f) d\mu$$

whenever g is a characteristic function. This identity now extends to non-negative simple functions by linearity and then to all non-negative measurable functions by MCT. The statements involving complex functions follows as in the solution to Exercise 8.5.

Remark 8.42. If X is a random variable on a probability space, $(\Omega, \mathcal{B}, P)$, and $F(x) := P(X \le x)$. Then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\, dF(x) \tag{8.11}$$

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

$$d\mu_F(x) = F'(x)\, dm(x) \tag{8.12}$$

and Eq. (8.11) may be written as

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) F'(x)\, dm(x). \tag{8.13}$$

To verify Eq. (8.12) it suffices to observe, by the fundamental theorem of calculus, that

$$\mu_F((a,b]) = F(b) - F(a) = \int_a^b F'(x)\, dx = \int_{(a,b]} F' dm.$$

From this equation we may deduce that $\mu_F(A) = \int_A F' dm$ for all $A \in \mathcal{B}_{\mathbb{R}}$.

Exercise 8.7. 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,

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h dm \tag{8.14}$$

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 5.1 to conclude $d\nu = F' dm$ on all of $\mathcal{B}_{\mathbb{R}}$. To prove Eq. (8.14) apply Exercise 8.6 with $g = h \circ F$ and $f = F^{-1}$.

Solution to Exercise (8.7). 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 Proposition 5.7. By Exercise 8.6 with $g = h \circ F$ and $f = F^{-1}$, we find

$$\int_{\mathbb{R}} h \circ F \cdot F' dm = \int_{\mathbb{R}} h \circ F d\nu = \int_{\mathbb{R}} h \circ F d\left(F_*^{-1} m\right) = \int_{\mathbb{R}} (h \circ F) \circ F^{-1} dm$$
$$= \int_{\mathbb{R}} h dm.$$

This result is also valid for all $h \in L^1(m)$.

Lemma 8.43. *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

$$P(X \geq x) \leq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{8.15}$$

and[1]

$$\lim_{x \to \infty} \frac{P(X \geq x)}{\frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}} = 1. \tag{8.16}$$

Proof. We begin by observing that

$$P(X \geq x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \leq \int_x^\infty \frac{1}{\sqrt{2\pi}} \frac{y}{x} e^{-y^2/2} dy = -\frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-y^2/2} |_x^{-\infty}$$

from which Eq. (8.15) follows. To prove Eq. (8.16), let $\alpha > 1$, then

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

[1] See, Gordon, Robert D. 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. (Reviewer: Z. W. Birnbaum) 62.0X

Hence

$$\frac{P(X \geq x)}{\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \geq \frac{\int_x^{\alpha x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy}{\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \geq \frac{1}{\alpha}\left[\frac{e^{-x^2/2}-e^{-\alpha^2x^2/2}}{e^{-x^2/2}}\right] = \frac{1}{\alpha}\left[1-e^{-\left(\alpha^2-1\right)x^2/2}\right].$$

From this equation it follows that

$$\liminf_{x\to\infty}\frac{P(X \geq x)}{\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \geq \frac{1}{\alpha}.$$

Since $\alpha > 1$ was arbitrary, it follows that

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

Since Eq. (8.15) implies that

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

we are done.

Additional information: Suppose that we now take

$$\alpha = 1 + x^{-p} = \frac{1+x^p}{x^p}.$$

Then

$$\left(\alpha^2-1\right)x^2 = \left(x^{-2p}+2x^{-p}\right)x^2 = \left(x^{2-2p}+2x^{2-p}\right).$$

Hence if $p = 2-\delta$, we find

$$\left(\alpha^2-1\right)x^2 = \left(x^{2(-1+\delta)}+2x^{\delta}\right) \leq 3x^{\delta}$$

so that

$$1 \geq \frac{P(X \geq x)}{\frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \geq \frac{1}{1+x^{-(2-\delta)}}\left[1-e^{-3x^{\delta}/2}\right]$$

for x sufficiently large. ■

Example 8.44. Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. standard normal random variables. Then

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\alpha c_n}e^{-\alpha^2c_n^2/2}.$$

Now, suppose that we take c_n so that

$$e^{-c_n^2/2} = \frac{C}{n}$$

or equivalently,

$$c_n^2/2 = \ln(n/C)$$

or

$$c_n = \sqrt{2\ln(n) - 2\ln(C)}.$$

(We now take $C = 1$.) It then follows that

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\alpha\sqrt{2\ln(n)}} e^{-\alpha^2 \ln(n)} = \frac{1}{\alpha\sqrt{2\ln(n)}} \frac{1}{n^{-\alpha^2}}$$

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 7.35 shows

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln n}} = 1 \text{ a.s..}$$

8.5 Measurability on Complete Measure Spaces

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

Proposition 8.45. *Suppose that $(X, \mathcal{B}, \mu)$ is a complete measure space*[2] *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,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{8.17}$$

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. (8.17) shows that $1_E g$ is

[2] 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.

measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ∎

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0, 1, 2\}$, $\mathcal{B} = \{\{0\}, \{1, 2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

Lemma 8.46. *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*

$$\int_X f d\bar{\mu} = \int_X g d\mu.$$

Proof. Suppose first that such a function g exists so that $\bar{\mu}(E) = 0$. Since g is also $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, we see from Proposition 8.45 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \backslash B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (8.18) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{8.18}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (8.18) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 6.32 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 $(\operatorname{Im} g)_\pm$. ∎

8.6 Comparison of the Lebesgue and the Riemann Integral

For the rest of this chapter, let $-\infty < a < b < \infty$ and $f : [a,b] \to \mathbb{R}$ be a bounded function. A partition of $[a,b]$ is a finite subset $\pi \subset [a,b]$ containing $\{a,b\}$. To each partition

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{8.19}$$

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 8.47. *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 8.18.

Lemma 8.48. *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

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{8.20}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} gdm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{8.21}$$

and

$$\int_{[a,b]} Gdm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{8.22}$$

Notation 8.49 *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 8.50. *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. (8.20), then*

$$G(x) = H(x) \ge f(x) \ge h(x) = g(x) \quad \forall\ x \notin \pi := \cup_{k=1}^{\infty} \pi_k. \tag{8.23}$$

   (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

$$G(x) \ge H(x) \ge f(x) \ge h(x) \ge g(x)\ \forall\ x \notin \pi. \tag{8.24}$$

   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

$$H(x) = \limsup_{y \to x} f(y) \geq G(x)$$

for all $x \notin \pi$. Combining this equation with Eq. (8.24) 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. (8.23) 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 8.51. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{8.25}$$

and the following statements are equivalent:

1. $H(x) = h(x)$ for m -a.e. x,
2. the set

$$E := \{x \in [a,b] : f \text{ is discontinuous at } x\}$$

is an $\bar{m}$ – null set.
3. f is Riemann integrable.

If f is Riemann integrable then f is Lebesgue measurable[3], i.e. f is $\mathcal{L}/\mathcal{B}$ – measurable where $\mathcal{L}$ is the Lebesgue σ – algebra and $\mathcal{B}$ is the Borel σ – algebra on $[a,b]$. Moreover if we let $\bar{m}$ denote the completion of m, then

$$\int_{[a,b]} H dm = \int_a^b f(x) dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{8.26}$$

Proof. Let $\{\pi_k\}_{k=1}^{\infty}$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 8.48 and let G and g be defined as in Lemma 8.50. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (8.25) is a consequence of Eqs. (8.21) and (8.22). From Eq. (8.25), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \leq f \leq H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (8.23), the remaining assertions including Eq. (8.26) are now consequences of Lemma 8.46. ■

Notation 8.52 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

[3] f need not be Borel measurable.

8.7 Exercises

Exercise 8.8. 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 8.9 (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 8.10. 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 8.11. 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 8.12. 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 8.13 (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 8.14. 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 8.15. 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 8.16. 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 8.17. 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 8.18. Prove Lemma 8.48.

8.7.1 Laws of Large Numbers Exercises

For the rest of the problems of this section, let $(\Omega, \mathcal{B}, P)$ be a probability space, $\{X_n\}_{n=1}^{\infty}$ be a sequence if i.i.d. random variables, and $S_n := \sum_{k=1}^{n} X_k$. If $\mathbb{E}|X_n| = \mathbb{E}|X_1| < \infty$ let

$$\mu := \mathbb{E}X_n \text{ – be the mean of } X_n,$$

if $\mathbb{E}\left[|X_n|^2\right] = \mathbb{E}\left[|X_1|^2\right] < \infty$, let

$$\sigma^2 := \mathbb{E}\left[(X_n - \mu)^2\right] = \mathbb{E}\left[X_n^2\right] - \mu^2 \text{ – be the standard deviation of } X_n$$

and if $\mathbb{E}\left[|X_n|^4\right] < \infty$, let

$$\gamma := \mathbb{E}\left[|X_n - \mu|^4\right].$$

Exercise 8.19 (A simple form of the Weak Law of Large Numbers). Assume $\mathbb{E}\left[|X_1|^2\right] < \infty$. Show

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu,$$

$$\mathbb{E}\left(\frac{S_n}{n} - \mu\right)^2 = \frac{\sigma^2}{n}, \text{ and}$$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$

for all $\varepsilon > 0$ and $n \in \mathbb{N}$.

Exercise 8.20 (A simple form of the Strong Law of Large Numbers). Suppose now that $\mathbb{E}\left[|X_1|^4\right] < \infty$. Show for all $\varepsilon > 0$ and $n \in \mathbb{N}$ that

$$\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^4\right] = \frac{1}{n^4}\left(n\gamma + 3n(n-1)\sigma^4\right)$$
$$= \frac{1}{n^2}\left[n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4\right]$$

and use this along with Chebyshev's inequality to show

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \le \frac{n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4}{\varepsilon^4 n^2}.$$

Conclude from the last estimate and the first Borel Cantelli Lemma 8.22 that $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s.

9

Functional Forms of the $\pi - \lambda$ Theorem

Notation 9.1 *Let Ω be a set and $\mathbb{H}$ be a subset of the bounded real valued functions on $\mathbb{H}$. 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 $f \in \mathbb{H}$. Similarly we say that $\mathbb{H}$ 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.

Proposition 9.2. *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. ∎

Theorem 9.3 (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 monotone convergence. If $\mathbb{M}$ is* ***multiplicative system*** *(i.e. $\mathbb{M}$ is a subset of $\mathbb{H}$ which is closed under pointwise multiplication), then $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M})$ – measurable functions.*

Proof. Let

$$\mathcal{L} := \{A \subset \Omega : 1_A \in \mathbb{H}\}.$$

We then have $\Omega \in \mathcal{L}$ since $1_\Omega = 1 \in \mathbb{H}$, if $A, B \in \mathcal{L}$ with $A \subset B$ then $B \setminus A \in \mathcal{L}$ since $1_{B\setminus A} = 1_B - 1_A \in \mathbb{H}$, and if $A_n \in \mathcal{L}$ with $A_n \uparrow A$, then $A \in \mathcal{L}$ because $1_{A_n} \in \mathbb{H}$ and $1_{A_n} \uparrow 1_A \in \mathbb{H}$. Therefore $\mathcal{L}$ is λ – system.

Let $\varphi_n(x) = 0 \vee [(nx) \wedge 1]$ (see Figure 9.1 below) 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 4.23, we may find polynomial functions, $p_l(x)$ such that $p_l \to \varphi_n$ uniformly on $[-M, M]$ .Since p_l is a polynomial it is easily seen that $\prod_{i=1}^{k} p_l \circ (f_i - a_i) \in \mathbb{H}$. Moreover,

$$\prod_{i=1}^{k} p_l \circ (f_i - a_i) \to F_n \text{ uniformly as } l \to \infty,$$

from with it follows that $F_n \in \mathbb{H}$ for all n. Since,

$$F_n \uparrow \prod_{i=1}^{k} 1_{\{f_i > a_i\}} = 1_{\cap_{i=1}^{k}\{f_i > a_i\}}$$

it follows that $1_{\cap_{i=1}^{k}\{f_i > a_i\}} \in \mathbb{H}$ or equivalently that $\cap_{i=1}^{k}\{f_i > a_i\} \in \mathcal{L}$. Therefore $\mathcal{L}$ contains the π – system, $\mathcal{P}$, consisting of finite intersections of sets of the form, $\{f > a\}$ with $f \in \mathbb{M}$ and $a \in \mathbb{R}$.

As a consequence of the above paragraphs and the $\pi - \lambda$ Theorem 7.4, $\mathcal{L}$ 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, 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.
■

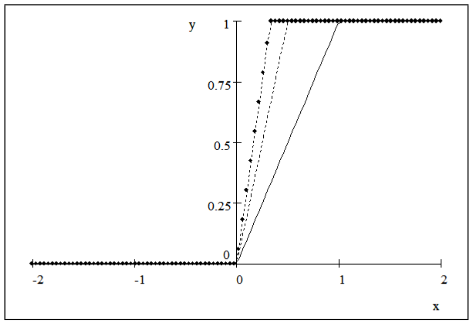


Fig. 9.1. Plots of φ_1, φ_2 and φ_3.

Corollary 9.4. *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* $\mathbb{M}$ *is a subset of* $\mathbb{H}$ *which is closed under pointwise multiplication, then* $\mathbb{H}$ *contains all bounded* $\sigma(\mathbb{M})$ *– measurable functions.*

Proof. This is of course a direct consequence of Theorem 9.3. Moreover, under the assumptions here, the proof of Theorem 9.3 simplifies in that Proposition 9.2 is no longer needed. For fun, let us give another self-contained proof of this corollary which does not even refer to the $\pi - \lambda$ theorem.

In this proof, we will 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.)

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 $f \in \mathbb{M}$, then $\mathbb{M} \subset \mathbb{H}^f$ and so 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 $fg \in \mathbb{H}$ whenever $f, g \in \mathbb{H}$, i.e. $\mathbb{H}$ is now an algebra of functions.

We will now show that $\mathcal{B} := \{A \subset \Omega : 1_A \in \mathbb{H}\}$ is 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.

Since $\mathbb{H}$ is a vector space, $\mathbb{H}$ 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, it follows that $\mathbb{H}$ contains all bounded $\mathcal{B}$ – measurable functions. The proof is now completed by showing $\mathcal{B}$ contains $\sigma(\mathbb{M})$ as was done in second paragraph of the proof of Theorem 9.3. ■

Exercise 9.1. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $X, Y : \Omega \to \mathbb{R}$ be a pair of random variables such that

$$\mathbb{E}\left[f(X)\, g(Y)\right] = \mathbb{E}\left[f(X)\, g(X)\right]$$

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

$$\mathbb{E}\left[h(X, Y)\right] = \mathbb{E}\left[h(X, X)\right].$$

Use Corollary 9.4 to show $\mathbb{H}$ is the vector space of all bounded Borel measurable functions. Then take $h(x, y) = 1_{\{x=y\}}$.

Corollary 9.5. *Suppose* $\mathbb{H}$ *is a real subspace of bounded functions such that* $1 \in \mathbb{H}$ *and* $\mathbb{H}$ *is closed under bounded convergence. If* $\mathcal{P} \subset 2^{\Omega}$ *is a multiplicative class such that* $1_A \in \mathbb{H}$ *for all* $A \in \mathcal{P}$, *then* $\mathbb{H}$ *contains all bounded* $\sigma(\mathcal{P})$ *– measurable functions.*

Proof. Let $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$. Then $\mathbb{M} \subset \mathbb{H}$ is a multiplicative system and the proof is completed with an application of Theorem 9.3. ■

Example 9.6. Suppose μ and ν are two probability measure on $(\Omega, \mathcal{B})$ such that

$$\int_{\Omega} f d\mu = \int_{\Omega} f d\nu \tag{9.1}$$

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 9.3 with $\mathbb{H}$ being the bounded measurable functions on Ω such that Eq. (9.1) holds. In particular if $\mathbb{M} = \{1\} \cup \{1_A : A \in \mathcal{P}\}$ with $\mathcal{P}$ being a multiplicative class we learn that $\mu = \nu$ on $\sigma(\mathbb{M}) = \sigma(\mathcal{P})$.

Corollary 9.7. *The smallest subspace of real valued functions,* $\mathbb{H}$, *on* $\mathbb{R}$ *which contains* $C_c(\mathbb{R}, \mathbb{R})$ *(the space of continuous functions on* $\mathbb{R}$ *with compact support) is the collection of bounded Borel measurable function on* $\mathbb{R}$.

Proof. By a homework problem, for $-\infty < a < b < \infty$, $1_{(a,b]}$ may be written as a bounded limit of continuous functions with compact support from which it follows that $\sigma(C_c(\mathbb{R}, \mathbb{R})) = \mathcal{B}_{\mathbb{R}}$. It is also easy to see that 1 is a bounded limit of functions in $C_c(\mathbb{R}, \mathbb{R})$ and hence $1 \in \mathbb{H}$. The corollary now follows by an application of The result now follows by an application of Theorem 9.3 with $\mathbb{M} := C_c(\mathbb{R}, \mathbb{R})$. ■

For the rest of this chapter, recall for $p \in [1, \infty)$ that $L^p(\mu) = L^p(X, \mathcal{B}, \mu)$ is the set of measurable functions $f : \Omega \to \mathbb{R}$ such that $\|f\|_{L^p} := \left(\int |f|^p \, d\mu\right)^{1/p} < \infty$. It is easy to see that $\|\lambda f\|_p = |\lambda| \, \|f\|_p$ for all $\lambda \in \mathbb{R}$ and we will show below that

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

i.e. $\|\cdot\|_p$ satisfies the triangle inequality.

Theorem 9.8 (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, $\lim_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} = 0$.[1] (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 $\|\varphi_n - \psi_k f_n\|_{L^p(\mu)} \le \frac{1}{n}$ then

$$\begin{aligned}\limsup_{n\to\infty} \|\psi_k f - \varphi_n\|_{L^p(\mu)} \le& \limsup_{n\to\infty} \|\psi_k (f - f_n)\|_{L^p(\mu)} \\ &+ \limsup_{n\to\infty} \|\psi_k f_n - \varphi_n\|_{L^p(\mu)} = 0 \qquad (9.2)\end{aligned}$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 9.3, now shows $\mathbb{H}$ contains all bounded measurable functions on Ω. Let $f \in L^p(\mu)$ be given. The dominated convergence theorem implies $\lim_{k\to\infty} \left\|\psi_k 1_{\{|f|\le k\}} f - f\right\|_{L^p(\mu)} = 0$. (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

$$\left\|\psi_k 1_{\{|f|\le k\}} f - \varphi_k\right\|_{L^p(\mu)} \le \frac{1}{k}.$$

The same line of reasoning used in Eq. (9.2) now implies $\lim_{k\to\infty} \|f - \varphi_k\|_{L^p(\mu)} = 0$. ∎

Example 9.9. 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 $1 \le p < \infty$. To see this, apply Theorem 9.8 with $\mathbb{M} = C_c(\mathbb{R}, \mathbb{R})$ and $\psi_k := 1_{[-k,k]}$.

Theorem 9.10. *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* $\mu(\{\varphi \ne 0\}) < \infty$. *Then* $\mathbb{S}(\mathcal{A}, \mu)$ *is dense subspace of* $L^p(\mu)$.

[1] It is at this point that the proof would break down if $p = \infty$.

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 9.8 after observing $\psi_k := 1_{\Omega_k} \in \mathbb{M}$ and $\psi_k \to 1$ boundedly. ∎

Theorem 9.11 (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 9.10. ∎

Notation 9.12 *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 9.13 (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 $\mu(h - f) < \varepsilon$.

Rework the Daniel integral section in the Analysis notes to stick to latticies of bounded functions.

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

$$\mu(\lambda h - \lambda f) = \lambda \mu(h - f) < \lambda \varepsilon.$$

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 9.3, to complete this proof, it suffices to show $\mathbb{H}$ is closed under monotone convergence. So suppose that $g_n \in \mathbb{H}$ and $g_n \uparrow g$, where $g : \Omega \to \mathbb{R}$ is a bounded function. Since $\mathbb{H}$ is a vector space, it follows that $0 \le \delta_n := g_{n+1} - g_n \in \mathbb{H}$ for all $n \in \mathbb{N}$. So if $\varepsilon > 0$ is given, we can find, $\mathbb{M}_\downarrow \ni u_n \le \delta_n \le v_n \in \mathbb{M}_\uparrow$ such that $\mu(v_n - u_n) \le 2^{-n}\varepsilon$ for all n. By replacing u_n by $u_n \vee 0 \in \mathbb{M}_\downarrow$ (by observation 1.), we may further assume that $u_n \ge 0$. Let

$$v := \sum_{n=1}^{\infty} v_n =\uparrow \lim_{N\to\infty} \sum_{n=1}^{N} v_n \in \mathbb{M}_\uparrow \text{ (using observations 2. and 5.)}$$

and for $N \in \mathbb{N}$, let

$$u^N := \sum_{n=1}^{N} u_n \in \mathbb{M}_\downarrow \text{ (using observation 2).}$$

Then

$$\sum_{n=1}^{\infty} \delta_n = \lim_{N\to\infty} \sum_{n=1}^{N} \delta_n = \lim_{N\to\infty} (g_{N+1} - g_1) = g - g_1$$

and $u^N \le g - g_1 \le v$. Moreover,

$$\mu\left(v - u^N\right) = \sum_{n=1}^{N} \mu(v_n - u_n) + \sum_{n=N+1}^{\infty} \mu(v_n) \le \sum_{n=1}^{N} \varepsilon 2^{-n} + \sum_{n=N+1}^{\infty} \mu(v_n)$$
$$\le \varepsilon + \sum_{n=N+1}^{\infty} \mu(v_n).$$

However, since

$$\sum_{n=1}^{\infty} \mu(v_n) \le \sum_{n=1}^{\infty} \mu\left(\delta_n + \varepsilon 2^{-n}\right) = \sum_{n=1}^{\infty} \mu(\delta_n) + \varepsilon\mu(\Omega)$$
$$= \sum_{n=1}^{\infty} \mu(g - g_1) + \varepsilon\mu(\Omega) < \infty,$$

it follows that for $N \in \mathbb{N}$ sufficiently large that $\sum_{n=N+1}^{\infty} \mu(v_n) < \varepsilon$. Therefore, for this N, we have $\mu\left(v - u^N\right) < 2\varepsilon$ and since $\varepsilon > 0$ is arbitrary, if follows that $g - g_1 \in \mathbb{H}$. Since $g_1 \in \mathbb{H}$ and $\mathbb{H}$ is a vector space, we may conclude that $g = (g - g_1) + g_1 \in \mathbb{H}$. ■

Theorem 9.14 (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 9.3, $\mathbb{H}^{\mathbb{R}}$ contains all bounded $\sigma\left(\mathbb{M}_0^{\mathbb{R}}\right)$ – 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 $\operatorname{Re} f = \frac{1}{2}\left(f + \bar{f}\right)$ and $\operatorname{Im} f = \frac{1}{2i}\left(f - \bar{f}\right)$ are in $\mathbb{H}$ or $\mathbb{M}_0$ respectively. Therefore $\mathbb{M}_0 = \mathbb{M}_0^{\mathbb{R}} + i\mathbb{M}_0^{\mathbb{R}}$, $\sigma\left(\mathbb{M}_0^{\mathbb{R}}\right) = \sigma\left(\mathbb{M}_0\right) = \sigma\left(\mathbb{M}\right)$, and $\mathbb{H} = \mathbb{H}^{\mathbb{R}} + i\mathbb{H}^{\mathbb{R}}$. Hence if $f : \Omega \to \mathbb{C}$ is a bounded $\sigma\left(\mathbb{M}\right)$ – measurable function, then $f = \operatorname{Re} f + i \operatorname{Im} f \in \mathbb{H}$ since $\operatorname{Re} f$ and $\operatorname{Im} f$ are in $\mathbb{H}^{\mathbb{R}}$. ■

10

Multiple and Iterated Integrals

10.1 Iterated Integrals

Notation 10.1 (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 10.2 *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.

10.2 Tonelli's Theorem and Product Measure

Theorem 10.3. *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* $y \in Y$,

$$x \to f(x,y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{10.1}$$

for each $x \in X$,

$$y \to f(x,y) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{10.2}$$

$$x \to \int_Y f(x,y)d\nu(y) \text{ is } \mathcal{M} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{10.3}$$

$$y \to \int_X f(x,y)d\mu(x) \text{ is } \mathcal{N} - \mathcal{B}_{[0,\infty]} \text{ measurable,} \tag{10.4}$$

and

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \int_Y d\nu(y) \int_X d\mu(x) f(x,y). \tag{10.5}$$

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. (10.1) and (10.2) 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. (10.3) holds and we have

$$\int_X d\mu(x) \int_Y d\nu(y) f(x,y) = \nu(B)\mu(A). \tag{10.6}$$

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. (10.4) and (10.5) 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. (10.1) – (10.5) 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 9.5 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. (10.1) – (10.5) 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. (10.1) –

(10.5) 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. (10.1) – (10.5) 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 10.4. *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*

$$\pi(E) = \int_X d\mu(x) \int_Y d\nu(y) 1_E(x, y) = \int_Y d\nu(y) \int_X d\mu(x) 1_E(x, y) \tag{10.7}$$

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 4.5 and 5.1 Proposition 4.26 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. (10.7) 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. (10.6). ■

Notation 10.5 *The measure π is called the product measure of μ and ν and will be denoted by $\mu \otimes \nu$.*

Theorem 10.6 (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

$$\int_{X \times Y} f \, d\pi = \int_X d\mu(x) \int_Y d\nu(y) f(x, y) \tag{10.8}$$

$$= \int_Y d\nu(y) \int_X d\mu(x) f(x, y). \tag{10.9}$$

Proof. By Theorem 10.3 and Corollary 10.4, 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 6.32, one deduces the theorem for general $f \in L^+(X \times Y, \mathcal{M} \otimes \mathcal{N})$. ■

Example 10.7. 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

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

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,

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

where $D_R = \left\{(x,y) : x^2 + y^2 < R^2\right\}$. Using the change of variables theorem described in Section 10.5 below,[1] we find

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

From this we learn that

$$I^2 = \lim_{R\to\infty} 2\pi \left(1 - e^{-R^2/2}\right) = 2\pi$$

as desired.

10.3 Fubini's Theorem

The following convention will be in force for the rest of this section.

Convention: If $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{C}$ is a measurable but non-integrable function, i.e. $\int_X |f|\, d\mu = \infty$, by convention we will define $\int_X f d\mu := 0$. However if f is a non-negative function (i.e. $f : X \to [0, \infty]$) is a non-integrable function we will still write $\int_X f d\mu = \infty$.

[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.

Theorem 10.8 (Fubini's Theorem). *Suppose* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *are* σ *– finite measure spaces,* $\pi = \mu \otimes \nu$ *is the product measure on* $\mathcal{M} \otimes \mathcal{N}$ *and* $f : X \times Y \to \mathbb{C}$ *is a* $\mathcal{M} \otimes \mathcal{N}$ *– measurable function. Then the following three conditions are equivalent:*

$$\int_{X \times Y} |f| \, d\pi < \infty, \text{ i.e. } f \in L^1(\pi), \tag{10.10}$$

$$\int_X \left(\int_Y |f(x,y)| \, d\nu(y) \right) d\mu(x) < \infty \text{ and} \tag{10.11}$$

$$\int_Y \left(\int_X |f(x,y)| \, d\mu(x) \right) d\nu(y) < \infty. \tag{10.12}$$

If any one (and hence all) of these condition hold, then $f(x,\cdot) \in L^1(\nu)$ *for* μ*-a.e.* x*,* $f(\cdot,y) \in L^1(\mu)$ *for* ν*-a.e.* y*,* $\int_Y f(\cdot,y) dv(y) \in L^1(\mu)$*,* $\int_X f(x,\cdot) d\mu(x) \in L^1(\nu)$ *and Eqs. (10.8) and (10.9) are still valid.*

Proof. The equivalence of Eqs. (10.10) – (10.12) is a direct consequence of Tonelli's Theorem 10.6. Now suppose $f \in L^1(\pi)$ is a real valued function and let

$$E := \left\{ x \in X : \int_Y |f(x,y)| \, d\nu(y) = \infty \right\}. \tag{10.13}$$

Then by Tonelli's theorem, $x \to \int_Y |f(x,y)| \, d\nu(y)$ is measurable and hence $E \in \mathcal{M}$. Moreover Tonelli's theorem implies

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

which implies that $\mu(E) = 0$. Let $f_\pm$ be the positive and negative parts of f, then using the above convention we have

$$\begin{aligned} \int_Y f(x,y) \, d\nu(y) &= \int_Y 1_{E^c}(x) f(x,y) \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) [f_+(x,y) - f_-(x,y)] \, d\nu(y) \\ &= \int_Y 1_{E^c}(x) f_+(x,y) \, d\nu(y) - \int_Y 1_{E^c}(x) f_-(x,y) \, d\nu(y). \end{aligned} \tag{10.14}$$

Noting that $1_{E^c}(x) f_\pm(x,y) = (1_{E^c} \otimes 1_Y \cdot f_\pm)(x,y)$ is a positive $\mathcal{M} \otimes \mathcal{N}$ – measurable function, it follows from another application of Tonelli's theorem that $x \to \int_Y f(x,y) \, d\nu(y)$ is $\mathcal{M}$ – measurable, being the difference of two measurable functions. Moreover

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

which shows $\int_Y f(\cdot,y)dv(y) \in L^1(\mu)$. Integrating Eq. (10.14) on x and using Tonelli's theorem repeatedly implies,

$$\begin{aligned}
&\int_X \left[\int_Y f(x,y)\,d\nu(y)\right] d\mu(x) \\
&\quad= \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x)\, f_+(x,y) - \int_X d\mu(x) \int_Y d\nu(y)\, 1_{E^c}(x)\, f_-(x,y) \\
&\quad= \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x)\, f_+(x,y) - \int_Y d\nu(y) \int_X d\mu(x)\, 1_{E^c}(x)\, f_-(x,y) \\
&\quad= \int_Y d\nu(y) \int_X d\mu(x)\, f_+(x,y) - \int_Y d\nu(y) \int_X d\mu(x)\, f_-(x,y) \\
&\quad= \int_{X\times Y} f_+ d\pi - \int_{X\times Y} f_- d\pi = \int_{X\times Y} (f_+ - f_-)\, d\pi = \int_{X\times Y} f d\pi \quad (10.15)
\end{aligned}$$

which proves Eq. (10.8) holds.

Now suppose that $f = u + iv$ is complex valued and again let E be as in Eq. (10.13). Just as above we still have $E \in \mathcal{M}$ and $\mu(E) = 0$. By our convention,

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

which is measurable in x by what we have just proved. Similarly one shows $\int_Y f(\cdot,y)\,d\nu(y) \in L^1(\mu)$ and Eq. (10.8) still holds by a computation similar to that done in Eq. (10.15). The assertions pertaining to Eq. (10.9) may be proved in the same way. ■

The previous theorems have obvious generalizations to products of any finite number of σ – finite measure spaces. For example the following theorem holds.

Theorem 10.9. *Suppose $\{(X_i, \mathcal{M}_i, \mu_i)\}_{i=1}^n$ are σ – finite measure spaces and $X := X_1 \times \cdots \times X_n$. Then there exists a unique measure, π, on $(X, \mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n)$ such that*

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

(This measure and its completion will be denoted by $\mu_1 \otimes \cdots \otimes \mu_n$.) If $f : X \to [0,\infty]$ is a $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ – measurable function then

$$\int_X f d\pi = \int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \ldots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)})\; f(x_1,\ldots,x_n) \quad (10.16)$$

where σ is any permutation of $\{1,2,\ldots,n\}$. This equation also holds for any $f \in L^1(\pi)$ and moreover, $f \in L^1(\pi)$ iff

$$\int_{X_{\sigma(1)}} d\mu_{\sigma(1)}(x_{\sigma(1)}) \dots \int_{X_{\sigma(n)}} d\mu_{\sigma(n)}(x_{\sigma(n)}) \ |f(x_1, \dots, x_n)| < \infty$$

for some (and hence all) permutations, σ.

This theorem can be proved by the same methods as in the two factor case, see Exercise 10.4. Alternatively, one can use the theorems already proved and induction on n, see Exercise 10.5 in this regard.

Proposition 10.10. *Suppose that $\{X_k\}_{k=1}^n$ are random variables on a probability space $(\Omega, \mathcal{B}, P)$ and $\mu_k = P \circ X_k^{-1}$ is the distribution for X_k for $k = 1, 2, \dots, n$, and $\pi := P \circ (X_1, \dots, X_n)^{-1}$ is the joint distribution of $(X_1, \dots, X_n)$. Then the following are equivalent,*

1. *$\{X_k\}_{k=1}^n$ are independent,*
2. *for all bounded measurable functions, $f : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,*

$$\mathbb{E}f(X_1, \dots, X_n) = \int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n), \text{ (taken in any order)} \tag{10.17}$$

and

3. *$\pi = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$.*

Proof. $(1 \implies 2)$ Suppose that $\{X_k\}_{k=1}^n$ are independent and let $\mathbb{H}$ denote the set of bounded measurable functions, $f : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that Eq. (10.17) holds. Then it is easily checked that $\mathbb{H}$ is a vector space which contains the constant functions and is closed under bounded convergence. Moreover, if $f = 1_{A_1 \times \dots \times A_n}$ where $A_i \in \mathcal{B}_{\mathbb{R}}$, we have

$$\begin{aligned}
\mathbb{E}f(X_1, \dots, X_n) &= P((X_1, \dots, X_n) \in A_1 \times \dots \times A_n) \\
&= \prod_{j=1}^n P(X_j \in A_j) = \prod_{j=1}^n \mu_j(A_j) \\
&= \int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n).
\end{aligned}$$

Therefore, $\mathbb{H}$ contains the multiplicative system, $\mathbb{M} := \{1_{A_1 \times \dots \times A_n} : A_i \in \mathcal{B}_{\mathbb{R}}\}$ and so by the multiplicative systems theorem, $\mathbb{H}$ contains all bounded $\sigma(\mathbb{M}) = \mathcal{B}_{\mathbb{R}^n}$ – measurable functions.

$(2 \implies 3)$ Let $A \in \mathcal{B}_{\mathbb{R}^n}$ and $f = 1_A$ in Eq. (10.17) to conclude that

$$\begin{aligned}
\pi(A) &= P((X_1, \dots, X_n) \in A) = \mathbb{E}1_A(X_1, \dots, X_n) \\
&= \int_{\mathbb{R}^n} 1_A(x_1, \dots, x_n)\, d\mu_1(x_1) \dots d\mu_n(x_n) = (\mu_1 \otimes \dots \otimes \mu_n)(A).
\end{aligned}$$

$(3 \implies 1)$ This follows from the identity,

$$P((X_1,\dots,X_n)\in A_1\times\cdots\times A_n)=\pi(A_1\times\cdots\times A_n)=\prod_{j=1}^n\mu_j(A_j)$$
$$=\prod_{j=1}^n P(X_j\in A_j),$$

which is valid for all $A_j\in\mathcal{B}_{\mathbb{R}}$. ■

Example 10.11 (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 $P(X=Y)=0$. To prove this, 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

$$\begin{aligned}P(X=Y)=\mathbb{E}\left[1_{\{X=Y\}}\right]&=\int_{\mathbb{R}^2}1_{\{x=y\}}d(\mu\otimes\nu)(x,y)\\&=\int_{\mathbb{R}}d\nu(y)\int_{\mathbb{R}}d\mu(x)\,1_{\{x=y\}}=\int_{\mathbb{R}}\mu(\{y\})\,d\nu(y)\\&=\int_{\mathbb{R}}0\;d\nu(y)=0.\end{aligned}$$

Example 10.12. In this example we will show

$$\lim_{M\to\infty}\int_0^M\frac{\sin x}{x}dx=\pi/2. \tag{10.18}$$

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 10.13. We have

$$\int_0^\infty\frac{\sin x}{x}e^{-\Lambda x}dx=\frac{1}{2}\pi-\arctan\Lambda\text{ for all }\Lambda>0 \tag{10.19}$$

and for$\Lambda, M \in [0,\infty)$,

$$\left|\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx - \frac{1}{2}\pi + \arctan \Lambda\right| \le C\frac{e^{-M\Lambda}}{M} \tag{10.20}$$

where $C = \max_{x\ge 0} \frac{1+x}{1+x^2} = \frac{1}{2\sqrt{2}-2} \cong 1.2$. In particular Eq. (10.18) is valid.

To verify these assertions, first notice that by the fundamental theorem of calculus,

$$|\sin x| = \left|\int_0^x \cos y dy\right| \le \left|\int_0^x |\cos y|\, dy\right| \le \left|\int_0^x 1 dy\right| = |x|$$

so $\left|\frac{\sin x}{x}\right| \le 1$ for all $x \ne 0$. Making use of the identity

$$\int_0^\infty e^{-tx} dt = 1/x$$

and Fubini's theorem,

$$\begin{aligned}
\int_0^M \frac{\sin x}{x} e^{-\Lambda x} dx &= \int_0^M dx \sin x\, e^{-\Lambda x} \int_0^\infty e^{-tx} dt \\
&= \int_0^\infty dt \int_0^M dx \sin x\, e^{-(\Lambda+t)x} \\
&= \int_0^\infty \frac{1 - (\cos M + (\Lambda+t)\sin M)\, e^{-M(\Lambda+t)}}{(\Lambda+t)^2+1} dt \\
&= \int_0^\infty \frac{1}{(\Lambda+t)^2+1} dt - \int_0^\infty \frac{\cos M + (\Lambda+t)\sin M}{(\Lambda+t)^2+1} e^{-M(\Lambda+t)} dt \\
&= \frac{1}{2}\pi - \arctan\Lambda - \varepsilon(M,\Lambda) \qquad (10.21)
\end{aligned}$$

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. (10.21) proves Eq. (10.20) from which Eq. (10.18) follows by taking $\Lambda \to \infty$ and Eq. (10.19) follows (using the dominated convergence theorem again) by letting $M \to \infty$.

Note: you may skip the rest of this chapter!

10.4 Fubini's Theorem and Completions

Notation 10.14 *Given $E \subset X \times Y$ and $x \in X$, let*

$$_xE := \{y \in Y : (x,y) \in E\}.$$

Similarly if $y \in Y$ is given let

$$E_y := \{x \in X : (x,y) \in E\}.$$

If $f : X \times Y \to \mathbb{C}$ is a function let $f_x = f(x,\cdot)$ and $f^y := f(\cdot,y)$ so that $f_x : Y \to \mathbb{C}$ and $f^y : X \to \mathbb{C}$.

Theorem 10.15. *Suppose $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ are* ***complete*** *σ – finite measure spaces. Let $(X\times Y,\mathcal{L},\lambda)$ be the completion of $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$. If f is $\mathcal{L}$ – measurable and (a) $f \geq 0$ or (b) $f \in L^1(\lambda)$ then f_x is $\mathcal{N}$ – measurable for μ a.e. x and f^y is $\mathcal{M}$ – measurable for ν a.e. y and in case (b) $f_x \in L^1(\nu)$ and $f^y \in L^1(\mu)$ for μ a.e. x and ν a.e. y respectively. Moreover,*

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

and

$$\int_{X\times Y} f d\lambda = \int_Y d\nu \int_X d\mu\, f = \int_X d\mu \int_Y d\nu\, f.$$

Proof. If $E \in \mathcal{M}\otimes\mathcal{N}$ is a $\mu\otimes\nu$ null set (i.e. $(\mu\otimes\nu)(E) = 0$), then

$$0 = (\mu\otimes\nu)(E) = \int_X \nu(_xE)d\mu(x) = \int_X \mu(E_y)d\nu(y).$$

This shows that

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

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 $|h(x,y)| \leq 1_E(x,y)$ 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

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

For general $f \in L^1(\lambda)$, we may choose $g \in L^1(\mathcal{M}\otimes\mathcal{N}, \mu\otimes\nu)$ such that $f(x,y) = g(x,y)$ for $\lambda-$ a.e. (x,y). Define $h := f - g$. Then $h = 0,\ \lambda-$ a.e. Hence by what we have just proved and Theorem 10.6 $f = g + h$ 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 10.6, 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

$$\int_Y d\nu(y)\int_X d\mu(x) f(x,y) = \int_{X\times Y} f(x,y)d\lambda(x,y).$$

■

10.5 Lebesgue Measure on $\mathbb{R}^d$ and the Change of Variables Theorem

Notation 10.16 *Let*

$$m^d := \overbrace{m\otimes\cdots\otimes m}^{d\ times} \text{ on } \mathcal{B}_{\mathbb{R}^d} = \overbrace{\mathcal{B}_\mathbb{R}\otimes\cdots\otimes\mathcal{B}_\mathbb{R}}^{d\ times}$$

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

Notation 10.18 *I will often be sloppy in the sequel and write m for m^d and dx for $dm(x) = dm^d(x)$, i.e.*

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

Hopefully the reader will understand the meaning from the context.

Theorem 10.19. *Lebesgue measure m^d is translation invariant. Moreover m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0,1]^d) = 1$.*

Proof. Let $A = J_1 \times \cdots \times J_d$ with $J_i \in \mathcal{B}_\mathbb{R}$ and $x \in \mathbb{R}^d$. Then

$$x + A = (x_1 + J_1) \times (x_2 + J_2) \times \cdots \times (x_d + J_d)$$

and therefore by translation invariance of m on $\mathcal{B}_\mathbb{R}$ we find that

$$m^d(x + A) = m(x_1 + J_1) \ldots m(x_d + J_d) = m(J_1) \ldots m(J_d) = m^d(A)$$

and hence $m^d(x + A) = m^d(A)$ for all $A \in \mathcal{B}_{\mathbb{R}^d}$ since it holds for A in a multiplicative system which generates $\mathcal{B}_{\mathbb{R}^d}$. From this fact we see that the measure $m^d(x + \cdot)$ and $m^d(\cdot)$ have the same null sets. Using this it is easily seen that $m(x + A) = m(A)$ for all $A \in \mathcal{L}_d$. The proof of the second assertion is Exercise 10.6. ∎

Exercise 10.1. 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 10.20 (Change of Variables Theorem). *Let $\Omega \subset_o \mathbb{R}^d$ be an open set and $T : \Omega \to T(\Omega) \subset_o \mathbb{R}^d$ be a C^1 – diffeomorphism,[2] see Figure 10.1. Then for any Borel measurable function, $f : T(\Omega) \to [0, \infty]$,*

$$\int_\Omega f(T(x))\,|\det T'(x)|\,dx = \int_{T(\Omega)} f(y)\,dy, \tag{10.22}$$

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

[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.

$$T'(x) = \begin{bmatrix} \partial_1 T_1(x) & \dots & \partial_d T_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 T_d(x) & \dots & \partial_d T_d(x) \end{bmatrix}, \tag{10.23}$$

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

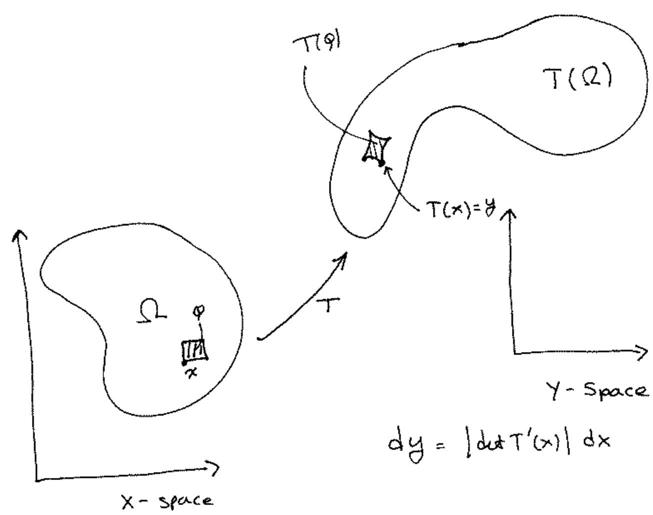


Fig. 10.1. The geometric setup of Theorem 10.20.

Remark 10.21. Theorem 10.20 is best remembered as the statement: if we make the change of variables $y = T(x)$, then $dy = |\det T'(x)|dx$. As usual, you must also change the limits of integration appropriately, i.e. if x ranges through Ω then y must range through $T(\Omega)$.

Proof. The proof will be by induction on d. The case $d = 1$ was essentially done in Exercise 8.7. 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

$$\int_\alpha^\beta |T'(x)|\,dx = -\int_\alpha^\beta T'(x)\,dx = T(\alpha) - T(\beta)$$
$$= m(T((\alpha,\beta])) = \int_{T([a,b])} 1_{T((\alpha,\beta])}(y)\,dy.$$

Combining the previous three equations shows

$$\int_{[a,b]} f(T(x))\,|T'(x)|\,dx = \int_{T([a,b])} f(y)\,dy \tag{10.24}$$

whenever f is of the form $f = 1_{T((\alpha,\beta])}$ with $a < \alpha < \beta < b$. An application of Dynkin's multiplicative system Theorem 9.3 then implies that Eq. (10.24) 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

$$\int_\Omega 1_{(a_n,b_n)} \cdot (f \circ T) \cdot |T'| dm = \int_\Omega \left(1_{T((a_n,b_n))} \cdot f\right) \circ T \cdot |T'| dm$$
$$= \lim_{k\to\infty} \int_\Omega \left(1_{T([\alpha_k,\beta_k])} \cdot f\right) \circ T \cdot |T'|\; dm$$
$$= \lim_{k\to\infty} \int_{T(\Omega)} 1_{T([\alpha_k,\beta_k])} \cdot f\; dm$$
$$= \int_{T(\Omega)} 1_{T((a_n,b_n))} \cdot f\; dm.$$

Summing this equality on n, then shows Eq. (10.22) 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. (10.23).

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

$$T(x) = (x_i, T_2(x), \dots, T_d(x)) \tag{10.25}$$

or

$$T(x) = (T_1(x), \dots, T_{d-1}(x), x_i) \tag{10.26}$$

for some $i \in \{1, \ldots, d\}$. For definiteness we will assume T is as in Eq. (10.25), the case of T in Eq. (10.26) may be handled similarly. For $t \in \mathbb{R}$, let $i_t : \mathbb{R}^{d-1} \to \mathbb{R}^d$ be the inclusion map defined by

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

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

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

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

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

see Figure 10.2. Expanding $\det T'(w_t)$ along the first row of the matrix $T'(w_t)$

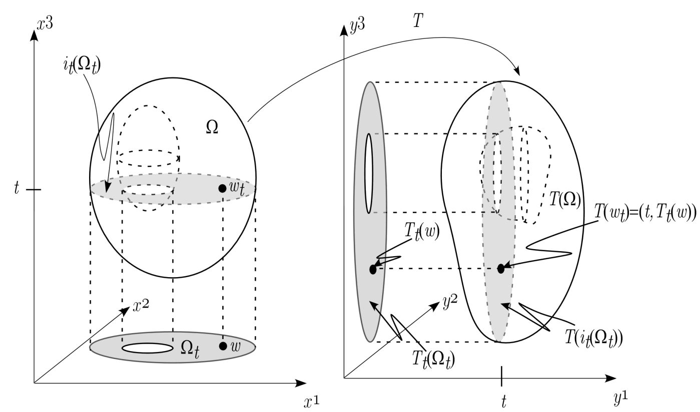


Fig. 10.2. In this picture $d = i = 3$ and Ω is an egg-shaped region with an egg-shaped hole. The picture indicates the geometry associated with the map T and slicing the set Ω along planes where $x_3 = t$.

shows

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

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

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

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

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

Case 2. (Eq. (10.22) 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 $T'(x_0)$, i.e. the determinant of $T'(x_0)$ with the first row and i^{th} – column removed. Since

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

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

$$S(x):=(x_i,T_2(x),\dots,T_d(x)). \tag{10.27}$$

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

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

Because

$$(T_1(x), \ldots, T_d(x)) = T(x) = R(S(x)) = R((x_i, T_2(x), \ldots, T_d(x)))$$

for all $x \in W$, if

$$(z_1, z_2, \ldots, z_d) = S(x) = (x_i, T_2(x), \ldots, T_d(x))$$

then

$$R(z) = \left(T_1\left(S^{-1}(z)\right), z_2, \ldots, z_d\right). \tag{10.28}$$

Observe that S is a map of the form in Eq. (10.25), R is a map of the form in Eq. (10.26), $T'(x) = R'(S(x))\, S'(x)$ (by the chain rule) and (by the multiplicative property of the determinant)

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

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

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

and Case 2. is proved.

Case 3. (General Case.) Let $f : \Omega \to [0, \infty]$ be a general non-negative Borel measurable function and let

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

Then each K_n is a compact subset of Ω and $K_n \uparrow \Omega$ as $n \to \infty$. Using the compactness of K_n and case 2, for each $n \in \mathbb{N}$, there is a finite open cover $\mathcal{W}_n$ of K_n such that $W \subset \Omega$ and Eq. (10.22) holds with Ω replaced by W for each $W \in \mathcal{W}_n$. Let $\{W_i\}_{i=1}^\infty$ be an enumeration of $\cup_{n=1}^\infty \mathcal{W}_n$ and set $\tilde{W}_1 = W_1$ and $\tilde{W}_i := W_i \setminus (W_1 \cup \cdots \cup W_{i-1})$ for all $i \geq 2$. Then $\Omega = \sum_{i=1}^\infty \tilde{W}_i$ and by repeated use of case 2.,

$$\begin{aligned}\int_\Omega f\circ T|\det T'|dm &= \sum_{i=1}^{\infty}\int_\Omega 1_{\tilde{W}_i}\cdot(f\circ T)\cdot|\det T'|dm \\ &= \sum_{i=1}^{\infty}\int_{W_i}\left[\left(1_{T(\tilde{W}_i)}f\right)\circ T\right]\cdot|\det T'|dm \\ &= \sum_{i=1}^{\infty}\int_{T(W_i)} 1_{T(\tilde{W}_i)}\cdot f\ dm = \sum_{i=1}^{n}\int_{T(\Omega)} 1_{T(\tilde{W}_i)}\cdot f\ dm \\ &= \int_{T(\Omega)} f dm.\end{aligned}$$

■

Remark 10.22. When $d=1$, one often learns the change of variables formula as

$$\int_a^b f(T(x))\,T'(x)\,dx = \int_{T(a)}^{T(b)} f(y)\,dy \qquad (10.29)$$

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. (10.29) is implies Eq. (10.22) with $\Omega=(a,b)$. On the other hand if $T'<0$ on (a,b) then $T((a,b))=(T(b),T(a))$ and Eq. (10.29) 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. (10.22). On the other hand Eq. (10.29) is more general than Eq. (10.22) since it does not require T to be injective. The standard proof of Eq. (10.29) 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,

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

An application of Dynkin's multiplicative systems theorem now shows that Eq. (10.29) holds for all bounded measurable functions f on (a,b). Then by the usual truncation argument, it also holds for all positive measurable functions on (a,b).

Example 10.23. Continuing the setup in Theorem 10.20, if $A \in \mathcal{B}_\Omega$, then

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

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

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

In particular if $T \in GL(d, \mathbb{R}) = GL(\mathbb{R}^d)$ – the space of $d \times d$ invertible matrices, then $m \circ T = |\det T|\, m$, i.e.

$$m(T(A)) = |\det T|\, m(A) \text{ for all} A \in \mathcal{B}_{\mathbb{R}^d}. \tag{10.30}$$

This equation also shows that $m \circ T$ and m have the same null sets and hence the equality in Eq. (10.30) is valid for any $A \in \mathcal{L}_d$.

Exercise 10.2. Show that $f \in L^1\left(T(\Omega), m^d\right)$ iff

$$\int_\Omega |f \circ T|\, |\det T'| dm < \infty$$

and if $f \in L^1\left(T(\Omega), m^d\right)$, then Eq. (10.22) holds.

Example 10.24 (Polar Coordinates). Suppose $T : (0,\infty) \times (0, 2\pi) \to \mathbb{R}^2$ is defined by

$$x = T(r,\theta) = (r\cos\theta, r\sin\theta),$$

i.e. we are making the change of variable,

$$x_1 = r\cos\theta \text{ and } x_2 = r\sin\theta \text{ for } 0 < r < \infty \text{ and } 0 < \theta < 2\pi.$$

In this case

$$T'(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and therefore

$$dx = \left|\det T'(r,\theta)\right| drd\theta = rdrd\theta.$$

Observing that

$$\mathbb{R}^2 \setminus T((0,\infty) \times (0,2\pi)) = \ell := \{(x,0) : x \geq 0\}$$

has m^2 – measure zero, it follows from the change of variables Theorem 10.20 that

$$\int_{\mathbb{R}^2} f(x)dx = \int_0^{2\pi} d\theta \int_0^\infty dr\ r \cdot f(r(\cos\theta, \sin\theta)) \tag{10.31}$$

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

Example 10.25 (Holomorphic Change of Variables). Suppose that $f : \Omega \subset_o \mathbb{C} \cong \mathbb{R}^2 \to \mathbb{C}$ is an injective holomorphic function such that $f'(z) \neq 0$ for all $z \in \Omega$. We may express f as

$$f(x+iy) = U(x,y) + iV(x,y)$$

for all $z = x + iy \in \Omega$. Hence if we make the change of variables,

$$w = u + iv = f(x+iy) = U(x,y) + iV(x,y)$$

then

$$dudv = \left| \det \begin{bmatrix} U_x & U_y \\ V_x & V_y \end{bmatrix} \right| dxdy = |U_x V_y - U_y V_x| \, dxdy.$$

Recalling that U and V satisfy the Cauchy Riemann equations, $U_x = V_y$ and $U_y = -V_x$ with $f' = U_x + iV_x$, we learn

$$U_x V_y - U_y V_x = U_x^2 + V_x^2 = |f'|^2 .$$

Therefore

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

Example 10.26. In this example we will evaluate the integral

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

where

$$\Omega = \left\{(x,y) : 1 < x^2 - y^2 < 2,\ 0 < xy < 1\right\},$$

see Figure 10.3. We are going to do this by making the change of variables,

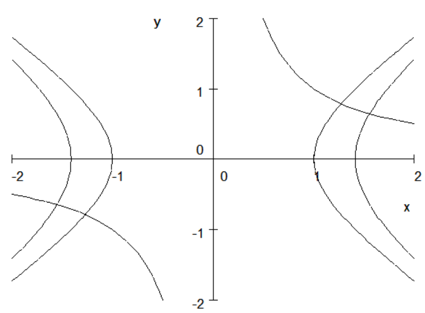


Fig. 10.3. The region Ω consists of the two curved rectangular regions shown.

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

in which case

$$dudv = \left|\det\begin{bmatrix} 2x & -2y \\ y & x \end{bmatrix}\right| dxdy = 2\left(x^2+y^2\right)dxdy$$

Notice that

$$\left(x^4 - y^4\right) = \left(x^2 - y^2\right)\left(x^2 + y^2\right) = u\left(x^2+y^2\right) = \frac{1}{2}ududv.$$

The function T is not injective on Ω but it is injective on each of its connected components. Let D be the connected component in the first quadrant so that $\Omega = -D \cup D$ and $T(\pm D) = (1,2)\times(0,1)$. The change of variables theorem then implies

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

and therefore $I = I_+ + I_- = 2\cdot(3/4) = 3/2.$

Exercise 10.3 (Spherical Coordinates). Let $T : (0,\infty)\times(0,\pi)\times(0,2\pi) \to \mathbb{R}^3$ be defined by

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

see Figure 10.4. By making the change of variables $x = T(r,\varphi,\theta)$, show

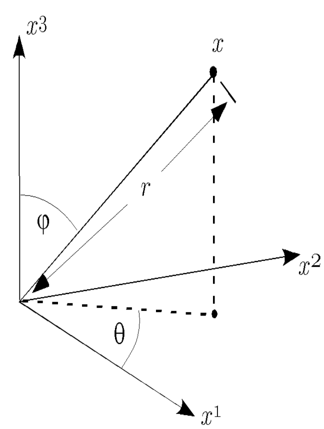


Fig. 10.4. The relation of x to (r,ϕ,θ) in spherical coordinates.

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

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

Lemma 10.27. *Let* $a > 0$ *and*

$$I_d(a) := \int\limits_{\mathbb{R}^d} e^{-a|x|^2} dm(x).$$

Then $I_d(a) = (\pi/a)^{d/2}$.

Proof. By Tonelli's theorem and induction,

$$\begin{aligned} I_d(a) &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-a|y|^2} e^{-at^2} m_{d-1}(dy)\ dt \\ &= I_{d-1}(a) I_1(a) = I_1^d(a). \end{aligned} \tag{10.32}$$

So it suffices to compute:

$$I_2(a) = \int\limits_{\mathbb{R}^2} e^{-a|x|^2} dm(x) = \int\limits_{\mathbb{R}^2\setminus\{0\}} e^{-a(x_1^2+x_2^2)} dx_1 dx_2.$$

Using polar coordinates, see Eq. (10.31), 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. (10.32). ■

10.6 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 10.28. *For $E \in \mathcal{B}_{S^{d-1}}$, let $\sigma(E) := d \cdot m(E_1)$. We call σ the surface measure on S^{d-1}.*

It is easy to check that σ is a measure. Indeed if $E \in \mathcal{B}_{S^{d-1}}$, then $E_1 = \Phi^{-1}((0,1] \times E) \in \mathcal{B}_{\mathbb{R}^d}$ so that $m(E_1)$ is well defined. Moreover if $E = \sum_{i=1}^{\infty} E_i$, then $E_1 = \sum_{i=1}^{\infty} (E_i)_1$ and

$$\sigma(E) = d \cdot m(E_1) = \sum_{i=1}^{\infty} m\left((E_i)_1\right) = \sum_{i=1}^{\infty} \sigma(E_i).$$

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 10.5 below. On the other hand

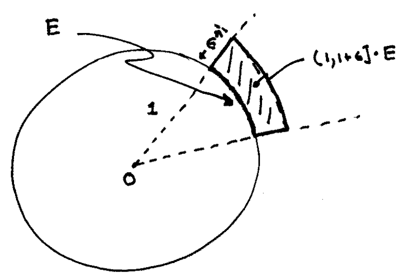


Fig. 10.5. Motivating the definition of surface measure for a sphere.

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

Therefore we expect the area of E should be given by

$$\sigma(E) = \lim_{\varepsilon \downarrow 0} \frac{\left\{(1+\varepsilon)^d - 1\right\} m(E_1)}{\varepsilon} = d \cdot m(E_1).$$

The following theorem is motivated by Example 10.24 and Exercise 10.3.

Theorem 10.29 (Polar Coordinates). *If $f : \mathbb{R}^d \to [0, \infty]$ is a $(\mathcal{B}_{R^d}, \mathcal{B})$–measurable function then*

$$\int_{\mathbb{R}^d} f(x) dm(x) = \int_{(0,\infty) \times S^{d-1}} f(r\omega) r^{d-1} \, dr d\sigma(\omega). \tag{10.33}$$

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

$$\int_{\mathbb{R}^d} f(|x|)dx = \int_0^\infty f(r)dV(r) \tag{10.34}$$

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 8.6,

$$\int_{\mathbb{R}^d} f dm = \int_{\mathbb{R}^d\setminus\{0\}} \left(f\circ\Phi^{-1}\right)\circ\Phi\; dm = \int_{(0,\infty)\times S^{d-1}} \left(f\circ\Phi^{-1}\right)\; d\left(\Phi_* m\right) \tag{10.35}$$

and therefore to prove Eq. (10.33) we must work out the measure $\Phi_* m$ on $\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$ defined by

$$\Phi_* m(A) := m\left(\Phi^{-1}(A)\right)\;\forall\; A\in\mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}. \tag{10.36}$$

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,

$$\begin{aligned}(\Phi_* m)\left((a,b]\times E\right) &= m\left(bE_1\setminus aE_1\right) = m(bE_1)-m(aE_1)\\ &= b^d m(E_1) - a^d m(E_1) = d\cdot m(E_1)\int_a^b r^{d-1}dr.\end{aligned} \tag{10.37}$$

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

$$\rho(J) = \int_J r^{d-1}dr\;\forall\; J\in\mathcal{B}_{(0,\infty)}, \tag{10.38}$$

Eq. (10.37) may be written as

$$(\Phi_* m)\left((a,b]\times E\right) = \rho((a,b])\cdot\sigma(E) = (\rho\otimes\sigma)\left((a,b]\times E\right). \tag{10.39}$$

Since

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

is a π class (in fact it is an elementary class) such that $\sigma(\mathcal{E}) = \mathcal{B}_{(0,\infty)}\otimes\mathcal{B}_{S^{d-1}}$, it follows from the $\pi-\lambda$ Theorem and Eq. (10.39) that $\Phi_* m = \rho\otimes\sigma$. Using this result in Eq. (10.35) gives

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

which combined with Tonelli's Theorem 10.6 proves Eq. (10.35). ■

Corollary 10.30. *The surface area $\sigma(S^{d-1})$ of the unit sphere $S^{d-1} \subset \mathbb{R}^d$ is*

$$\sigma(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)} \tag{10.40}$$

where Γ is the gamma function given by

$$\Gamma(x) := \int_0^\infty u^{x-1} e^{-u} du \tag{10.41}$$

Moreover, $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$.

Proof. Using Theorem 10.29 we find

$$I_d(1) = \int_0^\infty dr\, r^{d-1} e^{-r^2} \int\limits_{S^{d-1}} d\sigma = \sigma(S^{d-1}) \int_0^\infty r^{d-1} e^{-r^2} dr.$$

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

$$\begin{aligned} \int_0^\infty r^{d-1} e^{-r^2} dr &= \int_0^\infty u^{\frac{d-1}{2}} e^{-u} \frac{1}{2} u^{-1/2} du \\ &= \frac{1}{2} \int_0^\infty u^{\frac{d}{2}-1} e^{-u} du = \frac{1}{2} \Gamma(d/2). \end{aligned} \tag{10.42}$$

Combing the the last two equations with Lemma 10.27 which states that $I_d(1) = \pi^{d/2}$, we conclude that

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

which proves Eq. (10.40). Example 8.8 implies $\Gamma(1) = 1$ and from Eq. (10.42),

$$\begin{aligned} \Gamma(1/2) &= 2 \int_0^\infty e^{-r^2} dr = \int_{-\infty}^\infty e^{-r^2} dr \\ &= I_1(1) = \sqrt{\pi}. \end{aligned}$$

The relation, $\Gamma(x+1) = x\Gamma(x)$ is the consequence of the following integration by parts argument:

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

■

10.7 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 10.6, so that

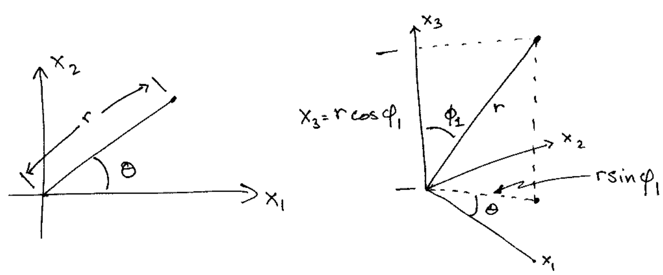


Fig. 10.6. Setting up polar coordinates in two and three dimensions.

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

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,

$$\begin{aligned} x_1 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\cos\theta \\ x_2 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\sin\varphi_1\sin\theta \\ x_3 &= r\sin\varphi_{n-2}\dots\sin\varphi_2\cos\varphi_1 \\ &\vdots \\ x_{n-2} &= r\sin\varphi_{n-2}\sin\varphi_{n-3}\cos\varphi_{n-4} \\ x_{n-1} &= r\sin\varphi_{n-2}\cos\varphi_{n-3} \\ x_n &= r\cos\varphi_{n-2}. \end{aligned} \tag{10.43}$$

By the change of variables formula,

$$\begin{aligned} &\int_{\mathbb{R}^n} f(x)dm(x) \\ &= \int_0^\infty dr \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} d\varphi_1\dots d\varphi_{n-2}d\theta \begin{bmatrix} \Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) \\ \times f(T_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)) \end{bmatrix} \end{aligned} \tag{10.44}$$

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 10.31. *The Jacobian,* Δ_n *is given by*

$$\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r) = r^{n-1}\sin^{n-2}\varphi_{n-2}\dots\sin^2\varphi_2\sin\varphi_1. \tag{10.45}$$

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

$$\begin{aligned} &\int_{rS^{n-1}} f(x)d\sigma(x) = r^{n-1}\int_{S^{n-1}} f(r\omega)d\sigma(\omega) \\ &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} f(T_n(\theta,\varphi_1,\dots,\varphi_{n-2},r))\Delta_n(\theta,\varphi_1,\dots,\varphi_{n-2},r)d\varphi_1\dots d\varphi_{n-2}d\theta \end{aligned} \tag{10.46}$$

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

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

i.e.

$$\Delta_{n+1}(\theta, \varphi_1, \ldots, \varphi_{n-2}, \varphi_{n-1}, r) = r\Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r\sin\varphi_{n-1}). \quad (10.47)$$

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 10.24, Eq. (10.47) implies,

$$\begin{aligned}
\Delta_3(\theta, \varphi_1, r) &= r\Delta_2(\theta, r\sin\varphi_1) = r^2 \sin\varphi_1 \\
\Delta_4(\theta, \varphi_1, \varphi_2, r) &= r\Delta_3(\theta, \varphi_1, r\sin\varphi_2) = r^3 \sin^2\varphi_2 \sin\varphi_1 \\
&\vdots \\
\Delta_n(\theta, \varphi_1, \ldots, \varphi_{n-2}, r) &= r^{n-1} \sin^{n-2}\varphi_{n-2} \ldots \sin^2\varphi_2 \sin\varphi_1
\end{aligned}$$

which proves Eq. (10.45). Equation (10.46) now follows from Eqs. (10.33), (10.44) and (10.45). ∎

As a simple application, Eq. (10.46) implies

$$\begin{aligned}
\sigma(S^{n-1}) &= \int_{0\le\varphi_i\le\pi, 0\le\theta\le 2\pi} \sin^{n-2}\varphi_{n-2} \ldots \sin^2\varphi_2 \sin\varphi_1 d\varphi_1 \ldots d\varphi_{n-2} d\theta \\
&= 2\pi \prod_{k=1}^{n-2} \gamma_k = \sigma(S^{n-2})\gamma_{n-2} \qquad (10.48)
\end{aligned}$$

where $\gamma_k := \int_0^\pi \sin^k \varphi d\varphi$. If $k \ge 1$, 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

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

Hence we may conclude

$$\gamma_0 = \pi, \; \gamma_1 = 2, \; \gamma_2 = \frac{1}{2}\pi, \; \gamma_3 = \frac{2}{3}2, \; \gamma_4 = \frac{3}{4}\frac{1}{2}\pi, \; \gamma_5 = \frac{4}{5}\frac{2}{3}2, \; \gamma_6 = \frac{5}{6}\frac{3}{4}\frac{1}{2}\pi$$

and more generally by induction that

$$\gamma_{2k} = \pi\frac{(2k-1)!!}{(2k)!!} \text{ and } \gamma_{2k+1} = 2\frac{(2k)!!}{(2k+1)!!}.$$

Indeed,

$$\gamma_{2(k+1)+1} = \frac{2k+2}{2k+3}\gamma_{2k+1} = \frac{2k+2}{2k+3}2\frac{(2k)!!}{(2k+1)!!} = 2\frac{[2(k+1)]!!}{(2(k+1)+1)!!}$$

and

$$\gamma_{2(k+1)} = \frac{2k+1}{2k+1}\gamma_{2k} = \frac{2k+1}{2k+2}\pi\frac{(2k-1)!!}{(2k)!!} = \pi\frac{(2k+1)!!}{(2k+2)!!}.$$

The recursion relation in Eq. (10.48) may be written as

$$\sigma(S^n) = \sigma\left(S^{n-1}\right)\gamma_{n-1} \tag{10.49}$$

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

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

and more generally that

$$\sigma(S^{2n}) = \frac{2\,(2\pi)^n}{(2n-1)!!} \text{ and } \sigma(S^{2n+1}) = \frac{(2\pi)^{n+1}}{(2n)!!} \tag{10.50}$$

which is verified inductively using Eq. (10.49). Indeed,

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

and

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

Using

$$(2n)!! = 2n\,(2(n-1))\ldots(2\cdot 1) = 2^n n!$$

we may write $\sigma(S^{2n+1}) = \frac{2\pi^{n+1}}{n!}$ which shows that Eqs. (10.33) and (10.50 are in agreement. We may also write the formula in Eq. (10.50) as

$$\sigma(S^n) = \begin{cases} \frac{2(2\pi)^{n/2}}{(n-1)!!} & \text{for } n \text{ even} \\ \frac{(2\pi)^{\frac{n+1}{2}}}{(n-1)!!} & \text{for } n \text{ odd.} \end{cases}$$

10.8 Exercises

Exercise 10.4. Prove Theorem 10.9. Suggestion, to get started define

$$\pi(A) := \int_{X_1} d\mu(x_1) \dots \int_{X_n} d\mu(x_n) 1_A(x_1, \dots, x_n)$$

and then show Eq. (10.16) holds. Use the case of two factors as the model of your proof.

Exercise 10.5. Let $(X_j, \mathcal{M}_j, \mu_j)$ for $j = 1, 2, 3$ be σ – finite measure spaces. Let $F : (X_1 \times X_2) \times X_3 \to X_1 \times X_2 \times X_3$ be defined by

$$F((x_1, x_2), x_3) = (x_1, x_2, x_3).$$

1. Show F is $((\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3, \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3)$ – measurable and F^{-1} is $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3)$ – measurable. That is
$$F : ((X_1 \times X_2)\times X_3, (\mathcal{M}_1 \otimes \mathcal{M}_2)\otimes\mathcal{M}_3) \to (X_1\times X_2\times X_3, \mathcal{M}_1\otimes\mathcal{M}_2\otimes\mathcal{M}_3)$$
is a "measure theoretic isomorphism."
2. Let $\pi := F_* [(\mu_1 \otimes \mu_2) \otimes \mu_3]$, i.e. $\pi(A) = [(\mu_1 \otimes \mu_2) \otimes \mu_3] (F^{-1}(A))$ for all $A \in \mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$. Then π is the unique measure on $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3$ such that
$$\pi(A_1 \times A_2 \times A_3) = \mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$$
for all $A_i \in \mathcal{M}_i$. We will write $\pi := \mu_1 \otimes \mu_2 \otimes \mu_3$.
3. Let $f : X_1 \times X_2 \times X_3 \to [0, \infty]$ be a $(\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function. Verify the identity,
$$\int_{X_1\times X_2\times X_3} f d\pi = \int_{X_3} d\mu_3(x_3) \int_{X_2} d\mu_2(x_2) \int_{X_1} d\mu_1(x_1) f(x_1, x_2, x_3),$$
makes sense and is correct.
4. (Optional.) Also show the above identity holds for any one of the six possible orderings of the iterated integrals.

Exercise 10.6. Prove the second assertion of Theorem 10.19. That is show m^d is the unique translation invariant measure on $\mathcal{B}_{\mathbb{R}^d}$ such that $m^d((0, 1]^d) = 1$. **Hint:** Look at the proof of Theorem 5.22.

Exercise 10.7. (Part of Folland Problem 2.46 on p. 69.) Let $X = [0, 1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$ be the Borel σ – field on X, m be Lebesgue measure on $[0, 1]$ and ν be counting measure, $\nu(A) = \#(A)$. Finally let $D = \{(x, x) \in X^2 : x \in X\}$ be the diagonal in X^2. Show

$$\int_X \left[\int_X 1_D(x, y)d\nu(y)\right] dm(x) \neq \int_X \left[\int_X 1_D(x, y)dm(x)\right] d\nu(y)$$

by explicitly computing both sides of this equation.

Exercise 10.8. Folland Problem 2.48 on p. 69. (Counter example related to Fubini Theorem involving counting measures.)

Exercise 10.9. Folland Problem 2.50 on p. 69 pertaining to area under a curve. (Note the $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$ should be $\mathcal{M} \otimes \mathcal{B}_{\bar{\mathbb{R}}}$ in this problem.)

Exercise 10.10. Folland Problem 2.55 on p. 77. (Explicit integrations.)

Exercise 10.11. Folland Problem 2.56 on p. 77. Let $f \in L^1((0,a), dm)$, $g(x) = \int_x^a \frac{f(t)}{t} dt$ for $x \in (0,a)$, show $g \in L^1((0,a), dm)$ and

$$\int_0^a g(x)dx = \int_0^a f(t)dt.$$

Exercise 10.12. Show $\int_0^\infty \left|\frac{\sin x}{x}\right| dm(x) = \infty$. So $\frac{\sin x}{x} \notin L^1([0,\infty), m)$ and $\int_0^\infty \frac{\sin x}{x} dm(x)$ is not defined as a Lebesgue integral.

Exercise 10.13. Folland Problem 2.57 on p. 77.

Exercise 10.14. Folland Problem 2.58 on p. 77.

Exercise 10.15. Folland Problem 2.60 on p. 77. Properties of the Γ – function.

Exercise 10.16. Folland Problem 2.61 on p. 77. Fractional integration.

Exercise 10.17. Folland Problem 2.62 on p. 80. Rotation invariance of surface measure on S^{n-1}.

Exercise 10.18. 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 10.19. Show, using Problem 10.17 that

$$\int_{S^{d-1}} \omega_i \omega_j d\sigma(\omega) = \frac{1}{d}\delta_{ij}\sigma\left(S^{d-1}\right).$$

Hint: show $\int_{S^{d-1}} \omega_i^2 d\sigma(\omega)$ is independent of i and therefore

$$\int_{S^{d-1}} \omega_i^2 d\sigma(\omega) = \frac{1}{d}\sum_{j=1}^d \int_{S^{d-1}} \omega_j^2 d\sigma(\omega).$$

11

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

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} \tag{11.1}$$

and when $p = \infty$, let

$$\|f\|_\infty = \inf \{a \geq 0 : \mu(|f| > a) = 0\} \tag{11.2}$$

For $0 < p \leq \infty$, 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 11.1. Suppose that $\|f\|_\infty \leq M$, 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. $|f(\omega)| \leq M$ for μ - a.e. ω. Conversely, if $|f| \leq M$ a.e. and $a > M$ then $\mu(|f| > a) = 0$ and hence $\|f\|_\infty \leq M$. This leads to the identity:

$$\|f\|_\infty = \inf \{a \geq 0 : |f(\omega)| \leq a \text{ for } \mu - \text{a.e. } \omega\}.$$

11.1 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 11.2. *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 11.3. *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 11.7 below.

Lemma 11.4 (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.*

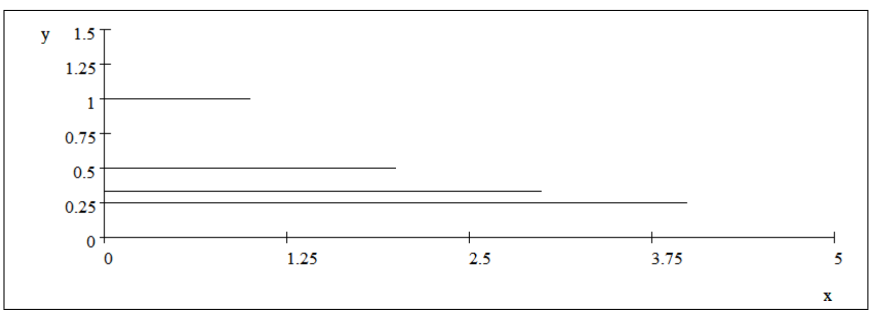
Proof. By Chebyshev's inequality (8.3),

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

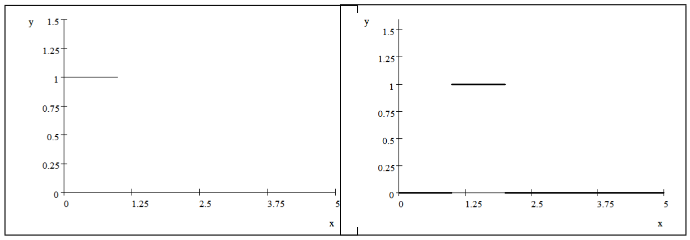
and therefore if $\{f_n\}$ is L^p – Cauchy, then

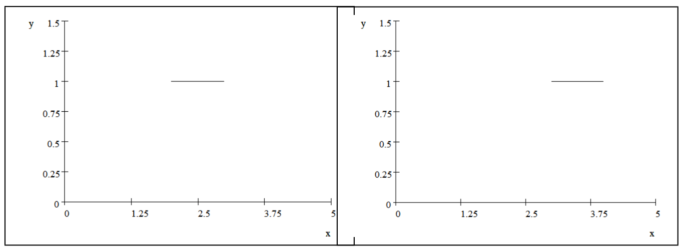
$$\mu\left(|f_n - f_m| \ge \varepsilon\right) \le \frac{1}{\varepsilon^p}\|f_n - f_m\|_p^p \to 0 \text{ as } m,n \to \infty$$

showing $\{f_n\}$ is L^0 – Cauchy. A similar argument holds for the L^p – convergent case. ■

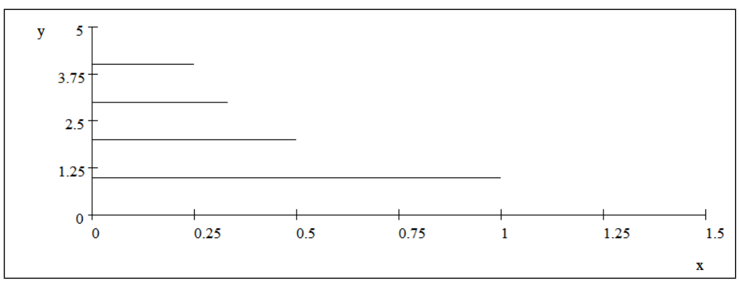


Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \nrightarrow 0$ in L^1, $f_n \xrightarrow{m} 0$.

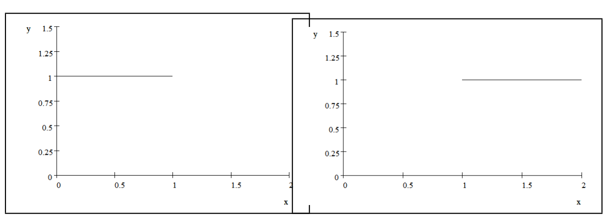


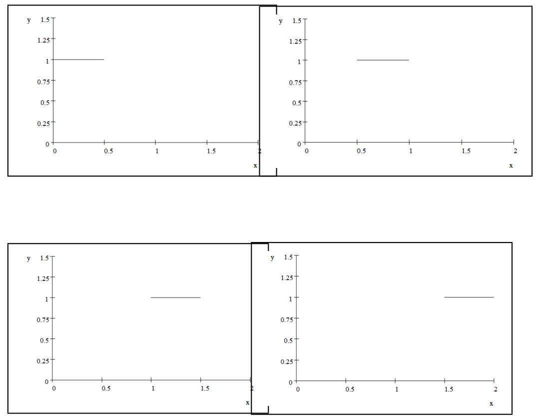


Above is a sequence of functions where $f_n \to 0$ a.e., yet $f_n \nrightarrow 0$ in L^1. or in measure.



Here is a sequence of functions where $f_n \to 0$ a.e., $f_n \xrightarrow{m} 0$ 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 11.5 (Egoroff's Theorem: almost sure convergence 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 $\varepsilon > 0$,

$$\begin{aligned} 0 &= \mu(\{|f_n - f| > \varepsilon \text{ i.o. } n\}) \\ &= \lim_{N\to\infty} \mu\left(\bigcup_{n\geq N} \{|f_n - f| > \varepsilon\}\right) \qquad (11.3) \\ &\geq \limsup_{N\to\infty} \mu\left(\{|f_N - f| > \varepsilon\}\right) \end{aligned}$$

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. (11.3) 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 11.6. *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

$$|a_m - a_n| = \left|\sum_{k=n}^{m-1}(a_{k+1} - a_k)\right| \le \sum_{k=n}^{m-1}|a_{k+1} - a_k| \le \sum_{k=n}^{\infty}\varepsilon_k := \delta_n\,. \tag{11.4}$$

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 (11.4) to find $|a - a_n| \le \delta_n$. ■

Theorem 11.7. *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 $\sum\limits_{n=1}^{\infty} \varepsilon_n < \infty$ ($\varepsilon_n = 2^{-n}$ would do) and set $\delta_n = \sum\limits_{k=n}^{\infty} \varepsilon_k$. 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 11.6, $f(\omega) := \lim\limits_{j\to\infty} g_j(\omega)$ 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 11.6 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,
$$\mu(|f - g_N| > \delta_N) \le \mu(F_N) \le \delta_N \to 0 \text{ as } N \to \infty$$
and in particular, $g_N \xrightarrow{\mu} f$ as $N \to \infty$.
With this in hand, it is straightforward to show $f_n \xrightarrow{\mu} f$. 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

$$\mu(\{|f_n - f| > \varepsilon\}) \le \mu(\{|f - g_j| > \varepsilon/2\}) + \mu(|g_j - f_n| > \varepsilon/2).$$

Therefore, letting $j \to \infty$ in this inequality gives,

$$\mu(\{|f_n - f| > \varepsilon\}) \le \limsup_{j\to\infty} \mu(|g_j - f_n| > \varepsilon/2) \to 0 \text{ as } n \to \infty$$

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 $\inf_k \mu(|f - f_{n_k}| \ge \varepsilon) > 0$. Any subsequence of $\{f_{n_k}\}$ would have the same property and hence can not be almost surely convergent because of Theorem 11.5.

■

Corollary 11.8 (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*

$$|f_n| \le g_n \text{ a.e., } f_n \xrightarrow{\mu} f, \; g_n \xrightarrow{\mu} g, \text{ and } \int g_n \to \int g \text{ as } n \to \infty.$$

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 $|f| \le g$ a.e. and hence $f \in L^1$ since $g \in L^1$. To see that $|f| \le g$, use Theorem 11.7 to find subsequences $\{f_{n_k}\}$ and $\{g_{n_k}\}$ of $\{f_n\}$ and $\{g_n\}$ respectively which are almost everywhere convergent. Then

$$|f| = \lim_{k\to\infty} |f_{n_k}| \le \lim_{k\to\infty} g_{n_k} = g \text{ a.e.}$$

If (for sake of contradiction) $\lim_{n\to\infty} \|f - f_n\|_1 \ne 0$ there exists $\varepsilon > 0$ and a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that

$$\int |f - f_{n_k}| \ge \varepsilon \text{ for all } k. \tag{11.5}$$

Using Theorem 11.7 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, $|f - f_{n_k}| \le g + g_{n_k} \to 2g$ and $\int (g + g_{n_k}) \to \int 2g$, an application of the dominated convergence Theorem 8.34 implies $\lim_{k\to\infty} \int |f - f_{n_k}| = 0$ which contradicts Eq. (11.5). ■

Exercise 11.1 (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 11.2. 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 $f_n \xrightarrow{\mu} f$ and $\int |f_n|^p \to \int |f|^p$.

Solution to Exercise (11.2). 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 $f_n \xrightarrow{\mu} f$ 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 $F_n \xrightarrow{\mu} 0$, $F_n \leq G_n \in L^1$, and $\int G_n \to \int G$ where $G := 2^p |f|^p \in L^1$. Therefore, by Corollary 11.8, $\int |f - f_n|^p = \int F_n \to \int 0 = 0$.

Corollary 11.9. *Suppose $(\Omega,\mathcal{B},\mu)$ is a probability space, $f_n \xrightarrow{\mu} f$ and $g_n \xrightarrow{\mu} g$ and $\varphi : \mathbb{R} \to \mathbb{R}$ and $\psi : \mathbb{R}^2 \to \mathbb{R}$ are continuous functions. Then*

1. $\varphi(f_n) \xrightarrow{\mu} \varphi(f)$,
2. $\psi(f_n, g_n) \xrightarrow{\mu} \psi(f,g)$,
3. $f_n + g_n \xrightarrow{\mu} f + g$, *and*
4. $f_n \cdot g_n \xrightarrow{\mu} f \cdot g$.

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 11.7.

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. ∎

11.2 Jensen's, Hölder's and Minikowski's Inequalities

Theorem 11.10 (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 7.31) and Figure 7.2 when φ is C^1 and Theorem 11.38 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 11.11. 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

$$\exp\left(\int_\Omega f d\mu\right) \leq \int_\Omega e^f d\mu, \tag{11.6}$$

$$\int_\Omega \log(|f|) d\mu \leq \log\left(\int_\Omega |f| \, d\mu\right)$$

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. (11.6), if $p_i, s_i > 0$ for $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \frac{1}{p_i} = 1$, then

$$s_1 \ldots s_n = e^{\sum_{i=1}^n \ln s_i} = e^{\sum_{i=1}^n \frac{1}{p_i} \ln s_i^{p_i}} \leq \sum_{i=1}^n \frac{1}{p_i} e^{\ln s_i^{p_i}} = \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}. \tag{11.7}$$

Indeed, we have applied Eq. (11.6) 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. (11.7), 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

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q. \tag{11.8}$$

(When $p = q = 1/2$, the inequality in Eq. (11.8) follows from the inequality, $0 \leq (s - t)^2$.)

As another special case of Eq. (11.7), take $p_i = n$ and $s_i = a_i^{1/n}$ with $a_i > 0$, then we get the arithmetic geometric mean inequality,

$$\sqrt[n]{a_1 \ldots a_n} \leq \frac{1}{n} \sum_{i=1}^n a_i. \tag{11.9}$$

Theorem 11.12 (Hölder's inequality). *Suppose that $1 \le p \le \infty$ and $q := \frac{p}{p-1}$, or equivalently $p^{-1} + q^{-1} = 1$. If f and g are measurable functions then*

$$\|fg\|_1 \le \|f\|_p \cdot \|g\|_q. \tag{11.10}$$

Assuming $p \in (1, \infty)$ and $\|f\|_p \cdot \|g\|_q < \infty$, equality holds in Eq. (11.10) iff $|f|^p$ and $|g|^q$ are linearly dependent as elements of L^1 which happens iff

$$|g|^q \|f\|_p^p = \|g\|_q^q |f|^p \quad a.e. \tag{11.11}$$

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. (11.10) 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. (11.8) gives,

$$\frac{|fg|}{\|f\|_p \|g\|_q} \le \frac{1}{p} \frac{|f|^p}{\|f\|_p} + \frac{1}{q} \frac{|g|^q}{\|g\|^q} \tag{11.12}$$

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. (11.12) implies

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \le \frac{1}{p} + \frac{1}{q} = 1$$

with equality iff Eq. (11.11) holds. The proof is finished since it is easily checked that equality holds in Eq. (11.10) when $|f|^p = c\,|g|^q$ of $|g|^q = c\,|f|^p$ for some constant c. ∎

Example 11.13. Suppose that $a_k \in \mathbb{C}$ for $k = 1, 2, \dots, n$ and $p \in [1, \infty)$, then

$$\left| \sum_{k=1}^n a_k \right|^p \le n^{p-1} \sum_{k=1}^n |a_k|^p. \tag{11.13}$$

Indeed, by Hölder's inequality applied using the measure space, $\{1, 2, \dots, n\}$ 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. (11.14).

Theorem 11.14 (Generalized Hölder's inequality). *Suppose that $f_i : \Omega \to \mathbb{C}$ are measurable functions for $i = 1, \dots, n$ and $p_1, \dots, p_n$ and r are positive numbers such that $\sum_{i=1}^n p_i^{-1} = r^{-1}$, then*

$$\left\| \prod_{i=1}^n f_i \right\|_r \le \prod_{i=1}^n \|f_i\|_{p_i}. \tag{11.14}$$

Proof. One may prove this theorem by induction based on Hölder's Theorem 11.12 above. Alternatively we may give a proof along the lines of the proof of Theorem 11.12 which is what we will do here.

Since Eq. (11.14) 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. (11.7) to find

$$s_1^r \dots s_n^r \le \sum_{i=1}^n \frac{(s_i^r)^{p_i/r}}{p_i/r} = r \sum_{i=1}^n \frac{s_i^{p_i}}{p_i}.$$

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 11.15 (Minkowski's Inequality). *If* $1 \le p \le \infty$ *and* $f, g \in L^p$ *then*

$$\|f + g\|_p \le \|f\|_p + \|g\|_p. \tag{11.15}$$

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

$$|f + g|^p = |f + g||f + g|^{p-1} \le (|f| + |g|)|f + g|^{p-1}$$

[1] In light of Example 11.13, the last 2^p in the above inequality may be replaced by 2^{p-1}.

and then applying Holder's inequality with $q = p/(p-1)$ gives

$$\int_\Omega |f+g|^p d\mu \le \int_\Omega |f|\ |f+g|^{p-1} d\mu + \int_\Omega |g|\ |f+g|^{p-1} d\mu \le (\|f\|_p + \|g\|_p) \|\ |f+g|^{p-1}\ \|_q, \tag{11.16}$$

where

$$\||f+g|^{p-1}\|_q^q = \int_\Omega (|f+g|^{p-1})^q d\mu = \int_\Omega |f+g|^p d\mu = \|f+g\|_p^p. \tag{11.17}$$

Combining Eqs. (11.16) and (11.17) implies

$$\|f+g\|_p^p \le \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} \tag{11.18}$$

Solving this inequality for $\|f+g\|_p$ gives Eq. (11.15). ■

11.3 Completeness of L^p – spaces

Theorem 11.16. *Let $\|\cdot\|_\infty$ be as defined in Eq. (11.2), 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 11.15, $\|\cdot\|_\infty$ 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 6.32. 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 11.17 (Completeness of $L^p(\mu)$). *For $1 \le p \le \infty$, $L^p(\mu)$ equipped with the L^p – norm, $\|\cdot\|_p$ (see Eq. (11.1)), is a Banach space.*

Proof. By Minkowski's Theorem 11.15, $\|\cdot\|_p$ 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 11.16.

Let $\{f_n\}_{n=1}^\infty \subset L^p(\mu)$ be a Cauchy sequence. By Chebyshev's inequality (Lemma 11.4), $\{f_n\}$ is L^0-Cauchy (i.e. Cauchy in measure) and by Theorem 11.7 there exists a subsequence $\{g_j\}$ of $\{f_n\}$ such that $g_j \to f$ a.e. By Fatou's Lemma,

$$\begin{aligned}\|g_j - f\|_p^p &= \int \lim_{k\to\infty}\inf |g_j - g_k|^p d\mu \le \lim_{k\to\infty}\inf \int |g_j - g_k|^p d\mu \\ &= \lim_{k\to\infty}\inf \|g_j - g_k\|_p^p \to 0 \text{ as } j \to \infty.\end{aligned}$$

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 12.5 for an important example of the use of this theorem.

11.4 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 $q \ge p$ as is shown in Corollary 11.18 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 $q \ge p$, see Corollary 11.23 below.

Corollary 11.18. *If $\mu(\Omega) < \infty$ and $0 < p < q \le \infty$, then $L^q(\mu) \subset L^p(\mu)$, the inclusion map is bounded and in fact*

$$\|f\|_p \le [\mu(\Omega)]^{\left(\frac{1}{p} - \frac{1}{q}\right)} \|f\|_q .$$

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 11.14,

$$\|f\|_p = \|f \cdot 1\|_p \le \|f\|_q \cdot \|1\|_a = \mu(\Omega)^{1/a}\|f\|_q = \mu(\Omega)^{\left(\frac{1}{p}-\frac{1}{q}\right)}\|f\|_q.$$

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 11.19 (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 11.18, $p \to \|a\|_p$ 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

$$\begin{aligned}\lim_{p\to 0}\|a\|_p &= \lim_{p\to 0}\left(\frac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p} = \lim_{p\to 0}\left(1 + p\frac{1}{n}\sum_{i=1}^n \ln a_i + O\left(p^2\right)\right)^{1/p} \\ &= e^{\frac{1}{n}\sum_{i=1}^n \ln a_i} = \sqrt[n]{a_1 \ldots a_n}.\end{aligned}$$

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 $p > 0$,

$$\frac{1}{n}M^p \le \frac{1}{n}\sum_{i=1}^n a_i^p \le M^p$$

and therefore,

$$\left(\frac{1}{n}\right)^{1/p} M \le \|a\|_p \le M.$$

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 11.20. *Suppose that $0 < p_0 < p_1 \le \infty$, $\lambda \in (0,1)$ and $p_\lambda \in (p_0,p_1)$ be defined by*

$$\frac{1}{p_\lambda} = \frac{1-\lambda}{p_0} + \frac{\lambda}{p_1} \tag{11.19}$$

with the interpretation that $\lambda/p_1 = 0$ if $p_1 = \infty$.[2] *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,

$$\begin{aligned}\|g\|_{p_0}^{p_0} &= \int |f|^{p_0} 1_{|f|>M} = M^{p_0}\int\left|\frac{f}{M}\right|^{p_0} 1_{|f|>M} \\ &\le M^{p_0}\int\left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|>M} \le M^{p_0-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty\end{aligned}$$

[2] A little algebra shows that λ may be computed in terms of p_0, p_λ and p_1 by

$$\lambda = \frac{p_0}{p_\lambda}\cdot\frac{p_1 - p_\lambda}{p_1 - p_0}.$$

and

$$\|h\|_{p_1}^{p_1} = \left\|f 1_{|f|\le M}\right\|_{p_1}^{p_1} = \int |f|^{p_1} 1_{|f|\le M} = M^{p_1}\int \left|\frac{f}{M}\right|^{p_1} 1_{|f|\le M}$$
$$\le M^{p_1}\int \left|\frac{f}{M}\right|^{p_\lambda} 1_{|f|\le M} \le M^{p_1-p_\lambda}\|f\|_{p_\lambda}^{p_\lambda} < \infty.$$

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 11.6 to the reader. ∎

Corollary 11.21 (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. (11.19), then $L^{p_0}\cap L^{p_1} \subset L^{p_\lambda}$ and*

$$\|f\|_{p_\lambda} \le \|f\|_{p_0}^{\lambda}\|f\|_{p_1}^{1-\lambda}. \tag{11.20}$$

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

$$\|f\|_{p_\lambda} \le \max\left(\lambda^{-1}, (1-\lambda)^{-1}\right)\left(\|f\|_{p_0} + \|f\|_{p_1}\right). \tag{11.21}$$

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 11.14,

$$\|f\|_{p_\lambda} = \left\||f|^{\lambda}|f|^{1-\lambda}\right\|_{p_\lambda} \le \left\||f|^{\lambda}\right\|_a \left\||f|^{1-\lambda}\right\|_b = \|f\|_{p_0}^{\lambda}\|f\|_{p_1}^{1-\lambda}.$$

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 11.4) $f_n \to f$ and $f_n \to g$ in measure and therefore by Theorem 11.7, $f = g$ a.e. It now is clear that $\lim_{n\to\infty}\|f - f_n\| = 0$. The estimate in Eq. (11.21) is left as Exercise 11.5 to the reader. ∎

Remark 11.22. Combining Proposition 11.20 and Corollary 11.21 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. (11.19).

Corollary 11.23. *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 11.21 with $p_0 = p$ and $p_1 = \infty$ to find

$$\|f\|_q \le \|f\|_p^{p/q} \|f\|_\infty^{1-p/q} \le \|f\|_p^{p/q} \|f\|_p^{1-p/q} = \|f\|_p.$$

■

11.4.1 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.

11.5 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 11.24 *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 11.25. *A collection of functions, $\Lambda \subset L^1(\mu)$ is said to be* ***uniformly integrable*** *if,*

$$\lim_{a \to \infty} \sup_{f \in \Lambda} \mu(|f| : |f| \geq a) = 0. \tag{11.22}$$

The condition in Eq. (11.22) implies $\sup_{f \in \Lambda} \|f\|_1 < \infty$.[3] Indeed, choose a sufficiently large so that $\sup_{f \in \Lambda} \mu(|f| : |f| \geq a) \leq 1$, then for $f \in \Lambda$

$$\|f\|_1 = \mu(|f| : |f| \geq a) + \mu(|f| : |f| < a) \leq 1 + a\mu(\Omega).$$

Let us also note that if $\Lambda = \{f\}$ with $f \in L^1(\mu)$, then Λ is uniformly integrable. Indeed, $\lim_{a \to \infty} \mu(|f| : |f| \geq a) = 0$ by the dominated convergence theorem.

Definition 11.26. *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*

$$\sup_{f \in \Lambda} \mu(|f| : E) < \varepsilon \text{ whenever } \mu(E) < \delta. \tag{11.23}$$

Remark 11.27. 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. (11.23) holds with $\varepsilon = 1$, then

$$\mu(|f_n|) = \sum_{\ell=1}^k \mu(|f_n| : E_\ell) \leq k$$

showing that $\mu(|f_n|) \leq k$ for all n.

Lemma 11.28 (This lemma may be skipped.). *For any $g \in L^1(\mu)$, $\Lambda = \{g\}$ is uniformly absolutely continuous.*

[3] 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 $L^1(m)$.

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 $\mu(|g| : E_n) \geq \varepsilon$ for all n. Since $|1_{E_n} g| \leq |g| \in L^1$ and for any $\delta > 0$, $\mu(1_{E_n} |g| > \delta) \leq \mu(E_n) \to 0$ as $n \to \infty$, the dominated convergence theorem of Corollary 11.8 implies $\lim_{n\to\infty} \mu(|g| : E_n) = 0$. This contradicts $\mu(|g| : E_n) \geq \varepsilon$ 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

$$\begin{aligned}\mu(|g| : E) &\leq \mu(|\varphi| : E) + \mu(|g - \varphi| : E) \\ &\leq \sum_{i=1}^n |c_i| \, \mu(E \cap B_i) + \|g - \varphi\|_1 \leq \left(\sum_{i=1}^n |c_i|\right) \mu(E) + \varepsilon/2.\end{aligned}$$

This shows $\mu(|g| : E) < \varepsilon$ provided that $\mu(E) < \varepsilon \left(2 \sum_{i=1}^n |c_i|\right)^{-1}$. ■

Proposition 11.29. *A subset $\Lambda \subset L^1(\mu)$ is uniformly integrable iff $\Lambda \subset L^1(\mu)$ is bounded is 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 $E \in \mathcal{B}$,

$$\begin{aligned}\mu(|f| : E) &= \mu(|f| : |f| \geq M, E) + \mu(|f| : |f| < M, E) \\ &\leq \sup_n \mu(|f| : |f| \geq M) + M\mu(E).\end{aligned}$$

So given $\varepsilon > 0$ choose M so large that $\sup_{f\in\Lambda} \mu(|f| : |f| \geq M) < \varepsilon/2$ 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

$$\mu(|f| \geq a) \leq \|f\|_1 / a \leq K/a \text{ for all } a > 0.$$

Hence given $\varepsilon > 0$ and $\delta > 0$ as in the definition of uniform absolute continuity, we may choose $a = K/\delta$ in which case

$$\sup_{f\in\Lambda} \mu(|f| : |f| \geq a) < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\lim_{a\to\infty} \sup_{f\in\Lambda} \mu(|f| : |f| \geq a) = 0$ as desired. ■

Corollary 11.30. *Suppose $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ are two uniformly integrable sequences, then $\{f_n + g_n\}_{n=1}^\infty$ is also uniformly integrable.*

Proof. By Proposition 11.29, $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ are both bounded in $L^1(\mu)$ and are both uniformly absolutely continuous. Since $\|f_n + g_n\|_1 \leq \|f_n\|_1 + \|g_n\|_1$ it follows that $\{f_n + g_n\}_{n=1}^\infty$ is bounded in $L^1(\mu)$ as well. Moreover, for $\varepsilon > 0$ we may choose $\delta > 0$ such that $\mu(|f_n| : E) < \varepsilon$ and $\mu(|g_n| : E) < \varepsilon$ whenever $\mu(E) < \delta$. For this choice of ε and δ, we then have

$$\mu(|f_n + g_n| : E) \leq \mu(|f_n| + |g_n| : E) < 2\varepsilon \text{ whenever } \mu(E) < \delta,$$

showing $\{f_n + g_n\}_{n=1}^\infty$ uniformly absolutely continuous. Another application of Proposition 11.29 completes the proof. ■

Exercise 11.3 (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 11.31 (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 11.1,

$$\int_\Omega |f|\, d\mu \le \liminf_{n\to\infty} \int_\Omega |f_n|\, d\mu \le M < \infty,$$

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 11.28 and Proposition 11.29) Λ_0 is uniformly absolutely continuous and hence Λ_0 is uniformly integrable. Therefore,

$$\begin{aligned} \|f - f_n\|_1 &= \mu(|f - f_n| : |f - f_n| \ge a) + \mu(|f - f_n| : |f - f_n| < a) \\ &\le \varepsilon(a) + \int_\Omega 1_{|f-f_n|<a} |f - f_n|\, d\mu \end{aligned} \tag{11.24}$$

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. (11.24), with the aid of the dominated convergence theorem (see Corollary 11.8), 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

$$\sup_n \mu(|f_n| : E) \le \sup_{n\le N} \mu(|f_n| : E) \vee (\varepsilon_N + \mu(|f| : E)) \le \varepsilon_N + \mu(g_N : E), \tag{11.25}$$

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 11.28) such that $\mu(g_N : E) < \varepsilon$ if $\mu(E) < \delta$. It then follows from Eq. (11.25) that

$$\sup_n \mu(|f_n| : E) < \varepsilon/2 + \varepsilon/2 = \varepsilon \text{ when } \mu(E) < \delta.$$

■

Example 11.32. 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

$$0 = \int_\Omega \lim_{\varepsilon \downarrow 0} f_\varepsilon dP \neq \lim_{\varepsilon \downarrow 0} \int_\Omega f_\varepsilon dP = 1.$$

This is a typical example of a bounded and pointwise convergent sequence in L^1 which is not uniformly integrable.

Example 11.33. 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 11.31, 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 11.34. *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 11.9, $|f_n|^p \xrightarrow{\mu} |f|^p$ in μ – measure, and $h_n := |f - f_n|^p \xrightarrow{\mu} 0$, and by Theorem 11.31, $|f|^p \in L^1(\mu)$ 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 11.8, 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 11.4. 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 11.8, $\lim_{n\to\infty} \int h_n \, d\mu = 0$, i.e. $|f_n|^p \to |f|^p$ in $L^1(\mu)$.[4] Hence it follows from Theorem 11.31 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 11.35. *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*
$$K := \sup_{f\in\Lambda} \mu(\varphi(|f|)) < \infty \tag{11.26}$$
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. (11.26) is valid.*

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 $\varepsilon_a := \sup_{x\ge a} \frac{x}{\varphi(x)} \to 0$ as $a \to \infty$ by assumption. Then for $f \in \Lambda$

[4] 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 $n \to \infty$.

$$\mu(|f| : |f| \geq a) = \mu\left(\frac{|f|}{\varphi(|f|)}\varphi(|f|) : |f| \geq a\right) \leq \mu(\varphi(|f|) : |f| \geq a)\varepsilon_a$$
$$\leq \mu(\varphi(|f|))\varepsilon_a \leq K\varepsilon_a$$

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.

$$\varphi(x) = \int_0^x \varphi'(y)dy = \sum_{n=0}^{\infty} (n+1)\,(x \wedge a_{n+1} - x \wedge a_n)\,.$$

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 $f \in \Lambda$,

$$\begin{aligned}\mu\left(\varphi(|f|)\right) &= \sum_{n=0}^{\infty} \mu\left(\varphi(|f|)1_{(a_n,a_{n+1}]}(|f|)\right)\\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{(a_n,a_{n+1}]}(|f|)\right)\\ &\leq \sum_{n=0}^{\infty} (n+1)\,\mu\left(|f|\, 1_{|f|\geq a_n}\right) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n}\end{aligned}$$

and hence

$$\sup_{f\in\Lambda} \mu\left(\varphi(|f|)\right) \leq \sum_{n=0}^{\infty} (n+1)\,\varepsilon_{a_n} < \infty.$$

■

11.6 Exercises

Exercise 11.4. 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 11.21 to show $\limsup_{q\to\infty} \|f\|_q \le \|f\|_\infty$ and to show $\liminf_{q\to\infty} \|f\|_q \ge \|f\|_\infty$, let $M < \|f\|_\infty$ and make use of Chebyshev's inequality.

Exercise 11.5. Prove Eq. (11.21) in Corollary 11.21. (Part of Folland 6.3 on p. 186.) **Hint:** Use the inequality, with $a, b \ge 1$ with $a^{-1} + b^{-1} = 1$ chosen appropriately,

$$st \le \frac{s^a}{a} + \frac{t^b}{b}$$

applied to the right side of Eq. (11.20).

Exercise 11.6. Complete the proof of Proposition 11.20 by showing $(L^p + L^r, \|\cdot\|)$ is a Banach space.

11.7 Appendix: Convex Functions

Reference; see the appendix (page 500) of Revuz and Yor.

Definition 11.36. *A function $\varphi : (a, b) \to \mathbb{R}$ is convex if for all $a < x_0 < x_1 < b$ and $t \in [0, 1]$ $\varphi(x_t) \le t\varphi(x_1) + (1-t)\varphi(x_0)$ where $x_t = tx_1 + (1-t)x_0$, see Figure **??** below.*

Example 11.37. The functions $\exp(x)$ and $-\log(x)$ are convex and $|x|^p$ is convex iff $p \ge 1$ as follows from Lemma 7.31 for $p > 1$ and by inspection of $p = 1$.

Theorem 11.38. *Suppose that $\varphi : (a, b) \to \mathbb{R}$ is convex and for $x, y \in (a, b)$ with $x < y$, let*[5]

$$F(x, y) := \frac{\varphi(y) - \varphi(x)}{y - x}.$$

Then;

1. *$F(x, y)$ is increasing in each of its arguments.*
2. *The following limits exist,*

$$\varphi'_+(x) := F(x, x+) := \lim_{y \downarrow x} F(x, y) < \infty \text{ and} \tag{11.27}$$

$$\varphi'_-(y) := F(y-, y) := \lim_{x \uparrow y} F(x, y) > -\infty. \tag{11.28}$$

[5] 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.

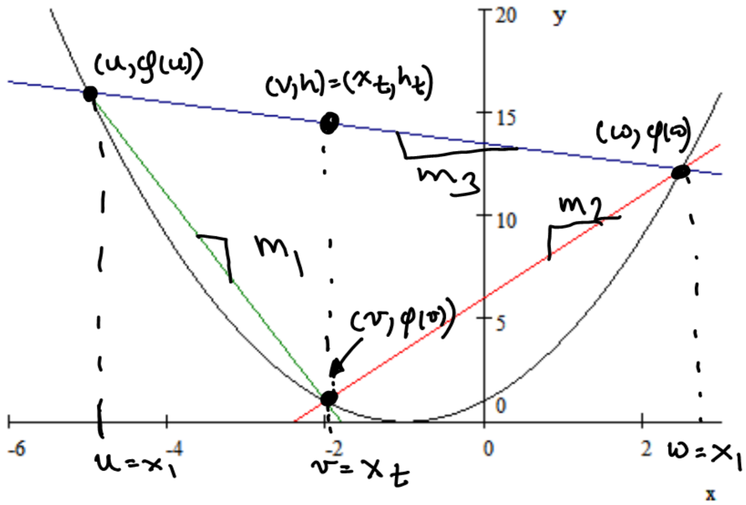


Fig. 11.1. A convex function with three cords. Notice the slope relationships; $m_1 \leq m_3 \leq m_2$.

3. *The functions, $\varphi'_{\pm}$ are both increasing functions and further satisfy,*

$$-\infty < \varphi'_{-}(x) \leq \varphi'_{+}(x) \leq \varphi'_{-}(y) < \infty \;\forall\; a < x < y < b. \tag{11.29}$$

4. *For any $t \in \left[\varphi'_{-}(x), \varphi'_{+}(x)\right]$,*

$$\varphi(y) \geq \varphi(x) + t(y - x) \;\text{ for all } x, y \in (a, b). \tag{11.30}$$

5. *For $a < \alpha < \beta < b$, let $K := \max\left\{\left|\varphi'_{+}(\alpha)\right|, \left|\varphi'_{-}(\beta)\right|\right\}$. Then*

$$|\varphi(y) - \varphi(x)| \leq K|y - x| \;\text{ for all } x, y \in [\alpha, \beta].$$

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. 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 $\varphi(x_t) \leq h_t$ for all $0 \leq t \leq 1$. 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

$$\frac{\varphi(x_1)-\varphi(x_0)}{x_1-x_0} = \frac{\varphi(x_1)-h_t}{x_1-x_t} \le \frac{\varphi(x_1)-\varphi(x_t)}{x_1-x_t} \text{ for all } x_0 \le x_t \le x_1$$

and 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,

$$\frac{\varphi(v)-\varphi(u)}{v-u} \le \frac{\varphi(w)-\varphi(u)}{w-u} \le \frac{\varphi(w)-\varphi(v)}{w-v}, \tag{11.31}$$

valid for all $a < u < v < w < b$, again see Figure 11.1. The first (second) inequality in Eq. (11.31) 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. (11.30) 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. (11.30) for all $x, y \in (a,b)$.

5. For $a < \alpha \le x < y \le \beta < b$, we have

$$\varphi'_+(\alpha) \le \varphi'_+(x) = F(x,x+) \le F(x,y) \le F(y-,y) = \varphi'_-(y) \le \varphi'_-(\beta) \tag{11.32}$$

and in particular,

$$-K \leq \varphi'_+(\alpha) \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \varphi'_-(\beta) \leq K.$$

This last inequality implies, $|\varphi(y) - \varphi(x)| \leq K(y - x)$ which is the desired Lipschitz bound.

6. For $a < c < x < y < b$, we have $\varphi'_+(x) = F(x, x+) \leq F(x, y)$ and letting $x \downarrow c$ (using the continuity of F) we learn $\varphi'_+(c+) \leq F(c, y)$. We may now let $y \downarrow c$ to conclude $\varphi'_+(c+) \leq \varphi'_+(c)$. Since $\varphi'_+(c) \leq \varphi'_+(c+)$, it follows that $\varphi'_+(c) = \varphi'_+(c+)$ and hence that φ'_+ is right continuous.

Similarly, for $a < x < y < c < b$, we have $\varphi'_-(y) \geq F(x, y)$ and letting $y \uparrow c$ (using the continuity of F) we learn $\varphi'_-(c-) \geq F(x, c)$. Now let $x \uparrow c$ to conclude $\varphi'_-(c-) \geq \varphi'_-(c)$. Since $\varphi'_-(c) \geq \varphi'_-(c-)$, 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. (11.29), 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 11.39. *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 $\psi_\pm(y) := \sup_{x \in D} [\varphi(x) + \varphi_\pm(x)(y - x)]$. According to Eq. (11.30) above, we know that $\varphi(y) \geq \psi_\pm(y)$ 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, $\psi_-(y) \geq \varphi(x_n) + \varphi'_-(x_n)(y - x_n)$, shows $\psi_-(y) \geq \varphi(x) + \varphi'_-(x)(y - x)$. Since $x \in (a, b)$ is arbitrary we may take $x = y$ to discover $\psi_-(y) \geq \varphi(y)$ and hence $\varphi(y) = \psi_-(y)$. The proof that $\varphi(y) = \psi_+(y)$ is similar. ■

Part III

Convergence Results

12

Laws of Large Numbers

In this chapter $\{X_k\}_{k=1}^{\infty}$ will be a sequence of random variables on a probability space, $(\Omega, \mathcal{B}, P)$, and we will set $S_n := X_1 + \cdots + X_n$ for all $n \in \mathbb{N}$.

Definition 12.1. *The* ***covariance****,* $\operatorname{Cov}(X,Y)$ *of two square integrable random variables,* X *and* Y*, is defined by*

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

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

$$\operatorname{Var}(X) := \operatorname{Cov}(X,X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 \tag{12.1}$$

We say that X *and* Y *are* ***uncorrelated*** *if* $\operatorname{Cov}(X,Y) = 0$*, 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* $\operatorname{Cov}(X_i, X_j) = 0$ *for all* $i \neq j$*.*

Notice that if X and Y are independent random variables, then $f(X)$, $g(Y)$ are independent and hence uncorrelated for any choice of Borel measurable functions, $f, g : \mathbb{R} \to \mathbb{R}$ such that $f(X)$ and $g(X)$ are square integrable. It also follows from Eq. (12.1) that

$$\operatorname{Var}(X) \leq \mathbb{E}\left[X^2\right] \text{ for all } X \in L^2(P). \tag{12.2}$$

Lemma 12.2. *The covariance function,* $\operatorname{Cov}(X,Y)$ *is bilinear in* X *and* Y *and* $\operatorname{Cov}(X,Y) = 0$ *if either* X *or* Y *is constant. For any constant* k*,* $\operatorname{Var}(X+k) = \operatorname{Var}(X)$ *and* $\operatorname{Var}(kX) = k^2 \operatorname{Var}(X)$*. If* $\{X_k\}_{k=1}^{n}$ *are uncorrelated* $L^2(P)$ *– random variables, then*

$$\operatorname{Var}(S_n) = \sum_{k=1}^{n} \operatorname{Var}(X_k).$$

Proof. We leave most of this simple proof to the reader. As an example of the type of argument involved, let us prove $\operatorname{Var}(X+k) = \operatorname{Var}(X)$;

$$\begin{aligned}\operatorname{Var}(X+k) &= \operatorname{Cov}(X+k, X+k) = \operatorname{Cov}(X+k, X) + \operatorname{Cov}(X+k, k) \\ &= \operatorname{Cov}(X+k, X) = \operatorname{Cov}(X,X) + \operatorname{Cov}(k, X) \\ &= \operatorname{Cov}(X,X) = \operatorname{Var}(X).\end{aligned}$$

■

Exercise 12.1 (A correlation inequality). Suppose that X is a random variable and $f, g : \mathbb{R} \to \mathbb{R}$ are two increasing functions such that both $f(X)$ and $g(X)$ are square integrable. Show $\operatorname{Cov}(f(X), g(X)) \geq 0$. **Hint:** let Y be another random variable which has the same law as X and is independent of X. Then consider

$$\mathbb{E}\left[(f(Y) - f(X)) \cdot (g(Y) - g(X))\right].$$

Theorem 12.3 (An L^2 – Weak Law of Large Numbers). *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of uncorrelated square integrable random variables, $\mu_n = \mathbb{E}X_n$ and $\sigma_n^2 = \operatorname{Var}(X_n)$. If there exists an increasing positive sequence, $\{a_n\}$ and $\mu \in \mathbb{R}$ such that*

$$\begin{aligned} &\frac{1}{a_n}\sum_{j=1}^{n} \mu_j \to \mu \text{ as } n \to \infty \text{ and} \\ &\frac{1}{a_n^2}\sum_{j=1}^{n} \sigma_j^2 \to 0 \text{ as } n \to \infty, \end{aligned}$$

then $\frac{S_n}{a_n} \to \mu$ in $L^2(P)$ and also in probability.

Proof. We first observe that $\mathbb{E}S_n = \sum_{j=1}^{n} \mu_j$ and

$$\mathbb{E}\left(S_n - \sum_{j=1}^{n} \mu_j\right)^2 = \operatorname{Var}(S_n) = \sum_{j=1}^{n} \operatorname{Var}(X_j) = \sum_{j=1}^{n} \sigma_j^2.$$

Hence

$$\mathbb{E}S_n = \frac{1}{a_n}\sum_{j=1}^{n} \mu_j \to \mu$$

and

$$\mathbb{E}\left(\frac{S_n - \sum_{j=1}^{n} \mu_j}{a_n}\right)^2 = \frac{1}{a_n^2}\sum_{j=1}^{n} \sigma_j^2 \to 0.$$

Hence,

$$\begin{aligned} \left\|\frac{S_n}{a_n} - \mu\right\|_{L^2(P)} &= \left\|\frac{S_n - \sum_{j=1}^{n} \mu_j}{a_n} + \frac{\sum_{j=1}^{n} \mu_j}{a_n} - \mu\right\|_{L^2(P)} \\ &\leq \left\|\frac{S_n - \sum_{j=1}^{n} \mu_j}{a_n}\right\|_{L^2(P)} + \left|\frac{\sum_{j=1}^{n} \mu_j}{a_n} - \mu\right| \to 0. \end{aligned}$$

■

Example 12.4. Suppose that $\{X_k\}_{k=1}^\infty \subset L^2(P)$ are uncorrelated identically distributed random variables. Then

$$\frac{S_n}{n} \overset{L^2(P)}{\to} \mu = \mathbb{E}X_1 \text{ as } n \to \infty.$$

To see this, simply apply Theorem 12.3 with $a_n = n$.

Proposition 12.5 (L^2 - Convergence of Random Sums). *Suppose that $\{X_k\}_{k=1}^\infty \subset L^2(P)$ are uncorrelated. If $\sum_{k=1}^\infty \operatorname{Var}(X_k) < \infty$ then*

$$\sum_{k=1}^\infty (X_k - \mu_k) \text{ converges in } L^2(P).$$

where $\mu_k := \mathbb{E}X_k$.

Proof. Letting $S_n := \sum_{k=1}^n (X_k - \mu_k)$, it suffices by the completeness of $L^2(P)$ (see Theorem 11.17) to show $\|S_n - S_m\|_2 \to 0$ as $m, n \to \infty$. Supposing $n > m$, we have

$$\begin{aligned}\|S_n - S_m\|_2^2 &= \mathbb{E}\left(\sum_{k=m+1}^n (X_k - \mu_k)\right)^2 \\ &= \sum_{k=m+1}^n \operatorname{Var}(X_k) = \sum_{k=m+1}^n \sigma_k^2 \to 0 \text{ as } m, n \to \infty.\end{aligned}$$

■

Note well: since $L^2(P)$ convergence implies $L^p(P)$ – convergence for $0 \le p \le 2$, where by $L^0(P)$ – **convergence** we mean convergence in probability. The remainder of this chapter is mostly devoted to proving *a.s.* convergence for the quantities in Theorem 11.17 and Proposition 12.5 under various assumptions. These results will be described in the next section.

12.1 Main Results

The proofs of most of the theorems in this section will be the subject of later parts of this chapter.

Theorem 12.6 (Khintchin's WLLN). *If $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^1(P)$ – random variables, then $\frac{1}{n}S_n \overset{P}{\to} \mu = \mathbb{E}X_1$.*

Proof. Letting

$$S_n' := \sum_{i=1}^n X_i 1_{|X_i| \le n},$$

we have $\{S_n' \ne S_n\} \subset \cup_{i=1}^n \{|X_i| > n\}$. Therefore, using Chebyschev's inequality along with the dominated convergence theorem, we have

$$
\begin{aligned}
P\left(S_n' \neq S_n\right) &\leq \sum_{i=1}^n P\left(|X_i| > n\right) = nP\left(|X_1| > n\right) \\
&\leq \mathbb{E}\left[|X_1| : |X_1| > n\right] \to 0.
\end{aligned}
$$

Hence it follows that

$$
P\left(\left|\frac{S_n}{n} - \frac{S_n'}{n}\right| > \varepsilon\right) \leq P\left(S_n' \neq S_n\right) \to 0 \text{ as } n \to \infty,
$$

i.e. $\frac{S_n}{n} - \frac{S_n'}{n} \xrightarrow{P} 0$. So it suffices to prove $\frac{S_n'}{n} \xrightarrow{P} \mu$.

We will now complete the proof by showing that, in fact, $\frac{S_n'}{n} \overset{L^2(P)}{\to} \mu$. To this end, let

$$
\mu_n := \frac{1}{n}\mathbb{E}S_n' = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[X_i 1_{|X_i| \leq n}\right] = \mathbb{E}\left[X_1 1_{|X_1| \leq n}\right]
$$

and observe that $\lim_{n\to\infty} \mu_n = \mu$ by the DCT. Moreover,

$$
\begin{aligned}
\mathbb{E}\left|\frac{S_n'}{n} - \mu_n\right|^2 &= \operatorname{Var}\left(\frac{S_n'}{n}\right) = \frac{1}{n^2}\operatorname{Var}\left(S_n'\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}\left(X_i 1_{|X_i| \leq n}\right) \\
&= \frac{1}{n}\operatorname{Var}\left(X_1 1_{|X_1| \leq n}\right) \leq \frac{1}{n}\mathbb{E}\left[X_1^2 1_{|X_1| \leq n}\right] \\
&\leq \mathbb{E}\left[|X_1| \, 1_{|X_1| \leq n}\right]
\end{aligned}
$$

and so again by the DCT, $\left\|\frac{S_n'}{n} - \mu_n\right\|_{L^2(P)} \to 0$. This completes the proof since,

$$
\left\|\frac{S_n'}{n} - \mu\right\|_{L^2(P)} \leq \left\|\frac{S_n'}{n} - \mu_n\right\|_{L^2(P)} + |\mu_n - \mu| \to 0 \text{ as } n \to \infty.
$$

■

In fact we have the stronger result.

Theorem 12.7 (Kolmogorov's Strong Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. Then there exists $\mu \in \mathbb{R}$ such that $\frac{1}{n}S_n \to \mu$ a.s. iff X_n is integrable and in which case $\mathbb{E}X_n = \mu$.*

Remark 12.8. If $\mathbb{E}|X_1| = \infty$ but $\mathbb{E}X_1^- < \infty$, then $\frac{1}{n}S_n \to \infty$ a.s. To prove this, for $M > 0$ let $X_n^M := X_n \wedge M$ and $S_n^M := \sum_{i=1}^n X_i^M$. It follows from Theorem 12.7 that $\frac{1}{n}S_n^M \to \mu^M := \mathbb{E}X_1^M$ a.s.. Since $S_n \geq S_n^M$, we may conclude that

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \liminf_{n\to\infty} \frac{1}{n} S_n^M = \mu^M \text{ a.s.}$$

Since $\mu^M \to \infty$ as $M \to \infty$, it follows that $\liminf_{n\to\infty} \frac{S_n}{n} = \infty$ a.s. and hence that $\lim_{n\to\infty} \frac{S_n}{n} = \infty$ a.s.

One proof of Theorem 12.7 is based on the study of random series. Theorem 12.11 and 12.12 are standard convergence criteria for random series.

Definition 12.9. *Two sequences, $\{X_n\}$ and $\{X_n'\}$, of random variables are **tail equivalent** if*

$$\mathbb{E}\left[\sum_{n=1}^{\infty} 1_{X_n \neq X_n'}\right] = \sum_{n=1}^{\infty} P\left(X_n \neq X_n'\right) < \infty.$$

Proposition 12.10. *Suppose $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Then*

1. *$\sum \left(X_n - X_n'\right)$ converges a.s.*
2. *The sum $\sum X_n$ is convergent a.s. iff the sum $\sum X_n'$ is convergent a.s. More generally we have*

$$P\left(\left\{\sum X_n \text{ is convergent}\right\} \triangle \left\{\sum X_n' \text{ is convergent}\right\}\right) = 1$$

3. *If there exists a random variable, X, and a sequence $a_n \uparrow \infty$ such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} X_k = X \text{ a.s}$$

then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} X_k' = X \text{ a.s}$$

Proof. If $\{X_n\}$ and $\{X_n'\}$ are tail equivalent, we know; for a.e. ω, $X_n(\omega) = X_n'(\omega)$ for a.a n. The proposition is an easy consequence of this observation. ■

Theorem 12.11 (Kolmogorov's Convergence Criteria). *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent square integrable random variables. If $\sum_{j=1}^{\infty} \operatorname{Var}(Y_j) < \infty$, then $\sum_{j=1}^{\infty} \left(Y_j - \mathbb{E}Y_j\right)$ converges a.s.*

Proof. One way to prove this is to appeal Proposition 12.5 above and Lévy's Theorem 12.31 below. As second method is to make use of Kolmogorov's inequality. We will give this second proof below. ■

The next theorem generalizes the previous theorem by giving necessary and sufficient conditions for a random series of independent random variables to converge.

Theorem 12.12 (Kolmogorov's Three Series Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are independent random variables. Then the random series, $\sum_{j=1}^\infty X_j$, is almost surely convergent iff there exists $c > 0$ such that*

1. $\sum_{n=1}^\infty P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$, *and*
3. $\sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ *converges.*

Moreover, if the three series above converge for some $c > 0$ then they converge for all values of $c > 0$.

Proof. Proof of sufficiency. Suppose the three series converge for some $c > 0$. If we let $X_n' := X_n 1_{|X_n| \le c}$, then

$$\sum_{n=1}^\infty P(X_n' \ne X_n) = \sum_{n=1}^\infty P(|X_n| > c) < \infty.$$

Hence $\{X_n\}$ and $\{X_n'\}$ are tail equivalent and so it suffices to show $\sum_{n=1}^\infty X_n'$ is almost surely convergent. However, by the convergence of the second series we learn

$$\sum_{n=1}^\infty \operatorname{Var}(X_n') = \sum_{n=1}^\infty \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$$

and so by Kolmogorov's convergence criteria,

$$\sum_{n=1}^\infty (X_n' - \mathbb{E}X_n') \text{ is almost surely convergent.}$$

Finally, the third series guarantees that $\sum_{n=1}^\infty \mathbb{E}X_n' = \sum_{n=1}^\infty \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ is convergent, therefore we may conclude $\sum_{n=1}^\infty X_n'$ is convergent. The proof of the reverse direction will be given in Section 12.8 below. ■

12.2 Examples

12.2.1 Random Series Examples

Example 12.13 (Kolmogorov's Convergence Criteria Example). Suppose that $\{Y_n\}_{n=1}^\infty$ are independent square integrable random variables, such that $\sum_{j=1}^\infty \operatorname{Var}(Y_j) < \infty$ and $\sum_{j=1}^\infty \mathbb{E}Y_j$ converges a.s., then $\sum_{j=1}^\infty Y_j$ converges a.s..

Definition 12.14. *A random variable, Y,* ***is normal with mean μ standard deviation σ^2*** *iff*

$$P(Y \in B) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_B e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ for all } B \in \mathcal{B}_{\mathbb{R}}. \tag{12.3}$$

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. (12.3) 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

$$\begin{aligned} \mathbb{E}[f(\sigma N + \mu)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\sigma x + \mu) e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \frac{dy}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} f(y) e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy. \end{aligned}$$

Lemma 12.15. *Suppose that* $\{Y_n\}_{n=1}^\infty$ *are independent square integrable random variables such that* $Y_n \overset{d}{=} N(\mu_n, \sigma_n^2)$. *Then* $\sum_{j=1}^\infty Y_j$ *converges a.s. iff* $\sum_{j=1}^\infty \sigma_j^2 < \infty$ *and* $\sum_{j=1}^\infty \mu_j$ *converges.*

Proof. The implication "$\implies$" is true without the assumption that the Y_n are normal random variables as pointed out in Example 12.13. To prove the converse directions we will make use of the Kolmogorov's three series theorem. Namely, if $\sum_{j=1}^\infty Y_j$ converges a.s. then the three series in Theorem 12.12 converge for all $c > 0$.

1. Since $Y_n \overset{d}{=} \sigma_n N + \mu_n$, we have for any $c > 0$ that

$$\infty > \sum_{n=1}^\infty P(|\sigma_n N + \mu_n| > c) = \sum_{n=1}^\infty \frac{1}{\sqrt{2\pi}} \int_{B_n} e^{-\frac{1}{2}x^2} dx \qquad (12.4)$$

where

$$B_n = (-\infty, -\frac{c+\mu_n}{\sigma_n}) \cup \left(\frac{c-\mu_n}{\sigma_n}, \infty\right).$$

If $\lim_{n\to\infty} \mu_n \neq 0$ then there is a $c > 0$ such that either $\mu_n \geq c$ i.o. or $\mu_n \leq -c$ i.o. In the first case in which case $(0,\infty) \subset B_n$ and in the second $(-\infty, 0) \subset B_n$ and in either case we will have $\frac{1}{\sqrt{2\pi}} \int_{B_n} e^{-\frac{1}{2}x^2} dx \geq 1/2$ i.o. which would contradict Eq. (12.4). Hence we may concluded that $\lim_{n\to\infty} \mu_n = 0$. Similarly if $\lim_{n\to\infty} \sigma_n \neq 0$, then we may conclude that B_n contains a set of the form $[\alpha, \infty)$ i.o. for some $\alpha < \infty$ and so

$$\frac{1}{\sqrt{2\pi}} \int_{B_n} e^{-\frac{1}{2}x^2} dx \geq \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-\frac{1}{2}x^2} dx \text{ i.o.}$$

which would again contradict Eq. (12.4). Therefore we may conclude that $\lim_{n\to\infty}\mu_n = \lim_{n\to\infty}\sigma_n = 0$.

2. The convergence of the second series for all $c > 0$ implies

$$\infty > \sum_{n=1}^{\infty} \operatorname{Var}\left(Y_n 1_{|Y_n|\le c}\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left(\left[\sigma_n N + \mu_n\right] 1_{|\sigma_n N+\mu_n|\le c}\right), \text{ i.e.}$$

$$\infty > \sum_{n=1}^{\infty} \left[\sigma_n^2 \operatorname{Var}\left(N 1_{|\sigma_n N+\mu_n|\le c}\right) + \mu_n^2 \operatorname{Var}\left(1_{|\sigma_n N+\mu_n|\le c}\right)\right] \ge \sum_{n=1}^{\infty} \sigma_n^2 \alpha_n.$$

where $\alpha_n := \operatorname{Var}\left(N 1_{|\sigma_n N+\mu_n|\le c}\right)$. As the reader should check, $\alpha_n \to 1$ as $n \to \infty$ and therefore we may conclude $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. It now follows by Kolmogorov's convergence criteria that $\sum_{n=1}^{\infty}(Y_n - \mu_n)$ is almost surely convergent and therefore

$$\sum_{n=1}^{\infty} \mu_n = \sum_{n=1}^{\infty} Y_n - \sum_{n=1}^{\infty} (Y_n - \mu_n)$$

converges as well.

Alternatively: we may also deduce the convergence of $\sum_{n=1}^{\infty} \mu_n$ by the third series as well. Indeed, for all $c > 0$ implies

$$\sum_{n=1}^{\infty} \mathbb{E}\left(\left[\sigma_n N + \mu_n\right] 1_{|\sigma_n N+\mu_n|\le c}\right) \text{ is convergent, i.e.}$$

$$\sum_{n=1}^{\infty} \left[\sigma_n \delta_n + \mu_n \beta_n\right] \text{ is convergent.}$$

where $\delta_n := \mathbb{E}\left(N \cdot 1_{|\sigma_n N+\mu_n|\le c}\right)$ and $\beta_n := \mathbb{E}\left(1_{|\sigma_n N+\mu_n|\le c}\right)$. With a little effort one can show,

$$\delta_n \sim e^{-k/\sigma_n^2} \text{ and } 1 - \beta_n \sim e^{-k/\sigma_n^2} \text{ for large } n.$$

Since $e^{-k/\sigma_n^2} \le C\sigma_n^2$ for large n, it follows that $\sum_{n=1}^{\infty} |\sigma_n \delta_n| \le C \sum_{n=1}^{\infty} \sigma_n^3 < \infty$ so that $\sum_{n=1}^{\infty} \mu_n \beta_n$ is convergent. Moreover,

$$\sum_{n=1}^{\infty} |\mu_n (\beta_n - 1)| \le C \sum_{n=1}^{\infty} |\mu_n| \sigma_n^2 < \infty$$

and hence

$$\sum_{n=1}^{\infty} \mu_n = \sum_{n=1}^{\infty} \mu_n \beta_n - \sum_{n=1}^{\infty} \mu_n (\beta_n - 1)$$

must also be convergent. ■

Example 12.16 (Brownian Motion). Let $\{N_n\}_{n=1}^{\infty}$ be i.i.d. standard normal random variable, i.e.

$$P(N_n \in A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \text{ for all } A \in \mathcal{B}_{\mathbb{R}}.$$

Let $\{\omega_n\}_{n=1}^{\infty} \subset \mathbb{R}$, $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}$, and $t \in \mathbb{R}$, then

$$\sum_{n=1}^{\infty} a_n N_n \sin \omega_n t \text{ converges a.s.}$$

provided $\sum_{n=1}^{\infty} a_n^2 < \infty$. This is a simple consequence of Kolmogorov's convergence criteria, Theorem 12.11, and the facts that $\mathbb{E}[a_n N_n \sin \omega_n t] = 0$ and

$$\operatorname{Var}(a_n N_n \sin \omega_n t) = a_n^2 \sin^2 \omega_n t \le a_n^2.$$

As a special case, if we take $\omega_n = (2n-1)\frac{\pi}{2}$ and $a_n = \frac{\sqrt{2}}{\pi(2n-1)}$, then it follows that

$$B_t := \frac{2\sqrt{2}}{\pi} \sum_{k=1,3,5,\dots} \frac{N_k}{k} \sin\left(k\frac{\pi}{2}t\right) \tag{12.5}$$

is a.s. convergent for all $t \in \mathbb{R}$. The factor $\frac{2\sqrt{2}}{\pi k}$ has been determined by requiring,

$$\int_0^1 \left[\frac{d}{dt} \frac{2\sqrt{2}}{\pi k} \sin(k\pi t)\right]^2 dt = 1$$

as seen by,

$$\begin{aligned}\int_0^1 \left[\frac{d}{dt} \sin\left(\frac{k\pi}{2}t\right)\right]^2 dt &= \frac{k^2\pi^2}{2^2} \int_0^1 \left[\cos\left(\frac{k\pi}{2}t\right)\right]^2 dt \\ &= \frac{k^2\pi^2}{2^2} \frac{2}{k\pi} \left[\frac{k\pi}{4}t + \frac{1}{4}\sin k\pi t\right]_0^1 = \frac{k^2\pi^2}{2^3}.\end{aligned}$$

Fact: Wiener in 1923 showed the series in Eq. (12.5) is in fact almost surely uniformly convergent. Given this, the process, $t \to B_t$ is almost surely continuous. The process $\{B_t : 0 \le t \le 1\}$ is **Brownian Motion.**

Example 12.17. As a simple application of Theorem 12.12, we will now use Theorem 12.12 to give a proof of Theorem 12.11. We will apply Theorem 12.12 with $X_n := Y_n - \mathbb{E}Y_n$. We need to then check the three series in the statement of Theorem 12.12 converge. For the first series we have by the Markov inequality,

$$\sum_{n=1}^{\infty} P(|X_n| > c) \le \sum_{n=1}^{\infty} \frac{1}{c^2} \mathbb{E}|X_n|^2 = \frac{1}{c^2} \sum_{n=1}^{\infty} \operatorname{Var}(Y_n) < \infty.$$

For the second series, observe that

$$\sum_{n=1}^{\infty} \mathrm{Var}\left(X_n 1_{|X_n|\le c}\right) \le \sum_{n=1}^{\infty} \mathbb{E}\left[\left(X_n 1_{|X_n|\le c}\right)^2\right] \le \sum_{n=1}^{\infty} \mathbb{E}\left[X_n^2\right] = \sum_{n=1}^{\infty} \mathrm{Var}\left(Y_n\right) < \infty$$

and for the third series (by Jensen's or Hölder's inequality)

$$\sum_{n=1}^{\infty} \left|\mathbb{E}\left(X_n 1_{|X_n|\le c}\right)\right| \le \sum_{n=1}^{\infty} \mathbb{E}\left(|X_n|^2 1_{|X_n|\le c}\right) \le \sum_{n=1}^{\infty} \mathrm{Var}\left(Y_n\right) < \infty.$$

12.2.2 A WLLN Example

Let $\{X_n\}_{n=1}^{\infty}$ be i.i.d. random variables with common distribution function, $F(x) := P(X_n \le x)$. For $x \in \mathbb{R}$ let $F_n(x)$ be the **empirical distribution function** defined by,

$$F_n(x) := \frac{1}{n}\sum_{j=1}^{n} 1_{X_j \le x} = \left(\frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}\right)((-\infty, x]).$$

Since $\mathbb{E}1_{X_j\le x} = F(x)$ and $\left\{1_{X_j\le x}\right\}_{j=1}^{\infty}$ are Bernoulli random variables, the weak law of large numbers implies $F_n(x) \overset{P}{\to} F(x)$ as $n \to \infty$. As usual, for $p \in (0,1)$ let

$$F^{\leftarrow}(p) := \inf\{x : F(x) \ge p\}$$

and recall that $F^{\leftarrow}(p) \le x$ iff $F(x) \ge p$. Let us notice that

$$F_n^{\leftarrow}(p) = \inf\{x : F_n(x) \ge p\} = \inf\left\{x : \sum_{j=1}^{n} 1_{X_j\le x} \ge np\right\}$$
$$= \inf\{x : \#\{j \le n : X_j \le x\} \ge np\}.$$

The **order statistic of** $(X_1, \dots, X_n)$ is the finite sequence, $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$, where $\left(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\right)$ denotes $(X_1, \dots, X_n)$ arranged in increasing order with possible repetitions. Let us observe that $X_k^{(n)}$ are all random variables for $k \le n$. Indeed, $X_k^{(n)} \le x$ iff $\#\{j \le n : X_j \le x\} \ge k$ iff $\sum_{j=1}^{n} 1_{X_j\le x} \ge k$, i.e.

$$\left\{X_k^{(n)} \le x\right\} = \left\{\sum_{j=1}^{n} 1_{X_j\le x} \ge k\right\} \in \mathcal{B}.$$

Moreover, if we let $\lceil x \rceil = \min\{n \in \mathbb{Z} : n \ge x\}$, the reader may easily check that $F_n^{\leftarrow}(p) = X_{\lceil np\rceil}^{(n)}$.

Proposition 12.18. *Keeping the notation above. Suppose that $p \in (0,1)$ is a point where*

$$F\left(F^{\leftarrow}(p)-\varepsilon\right) < p < F\left(F^{\leftarrow}(p)+\varepsilon\right) \text{ for all } \varepsilon > 0$$

then $X_{\lceil np \rceil}^{(n)} = F_n^{\leftarrow}(p) \xrightarrow{P} F^{\leftarrow}(p)$ as $n \to \infty$. Thus we can recover, with high probability, the p^{th} – quantile of the distribution F by observing $\{X_i\}_{i=1}^n$.

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned}\left\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\right\}^c &= \left\{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\right\} = \left\{F_n^{\leftarrow}(p) \le \varepsilon + F^{\leftarrow}(p)\right\} \\ &= \left\{F_n\left(\varepsilon + F^{\leftarrow}(p)\right) \ge p\right\}\end{aligned}$$

so that

$$\begin{aligned}\left\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\right\} &= \left\{F_n\left(F^{\leftarrow}(p)+\varepsilon\right) < p\right\} \\ &= \left\{F_n\left(\varepsilon + F^{\leftarrow}(p)\right) - F\left(\varepsilon + F^{\leftarrow}(p)\right) < p - F\left(F^{\leftarrow}(p)+\varepsilon\right)\right\}.\end{aligned}$$

Letting $\delta_\varepsilon := F\left(F^{\leftarrow}(p)+\varepsilon\right) - p > 0$, we have, as $n \to \infty$, that

$$P\left(\left\{F_n^{\leftarrow}(p) - F^{\leftarrow}(p) > \varepsilon\right\}\right) = P\left(F_n\left(\varepsilon + F^{\leftarrow}(p)\right) - F\left(\varepsilon + F^{\leftarrow}(p)\right) < -\delta_\varepsilon\right) \to 0.$$

Similarly, let $\delta_\varepsilon := p - F\left(F^{\leftarrow}(p)-\varepsilon\right) > 0$ and observe that

$$\left\{F^{\leftarrow}(p) - F_n^{\leftarrow}(p) \ge \varepsilon\right\} = \left\{F_n^{\leftarrow}(p) \le F^{\leftarrow}(p) - \varepsilon\right\} = \left\{F_n\left(F^{\leftarrow}(p)-\varepsilon\right) \ge p\right\}$$

and hence,

$$\begin{aligned}&P\left(F^{\leftarrow}(p) - F_n^{\leftarrow}(p) \ge \varepsilon\right) \\ &\quad = P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right) - F\left(F^{\leftarrow}(p)-\varepsilon\right) \ge p - F\left(F^{\leftarrow}(p)-\varepsilon\right)\right) \\ &\quad = P\left(F_n\left(F^{\leftarrow}(p)-\varepsilon\right) - F\left(F^{\leftarrow}(p)-\varepsilon\right) \ge \delta_\varepsilon\right) \to 0 \text{ as } n \to \infty.\end{aligned}$$

Thus we have shown that $X_{\lceil np \rceil}^{(n)} \xrightarrow{P} F^{\leftarrow}(p)$ as $n \to \infty$. ∎

12.3 Strong Law of Large Number Examples

Example 12.19 (Renewal Theory). Let $\{X_i\}_{i=1}^\infty$ be i.i.d. random variables with $0 < X_i < \infty$ a.s. Think of the X_i as the time that bulb number i burns and $T_n := X_1 + \cdots + X_n$ is the time that the n^{th} – bulb burns out. (We assume the bulbs are replaced immediately on burning out.) Further let $N_t := \sup\{n \ge 0 : T_n \le t\}$ denote the number of bulbs which have burned out up to time n. By convention, we set $T_0 = 0$. Letting $\mu := \mathbb{E}X_1 \in (0,\infty]$, we have $\mathbb{E}T_n = n\mu$ – the expected time the n^{th} – bulb burns out. On these grounds we expect $N_t \sim t/\mu$ and hence

$$\frac{1}{t}N_t \to \frac{1}{\mu} \text{ a.s.} \tag{12.6}$$

To prove Eq. (12.6), by the SSLN, if $\Omega_0 := \left\{\lim_{n\to\infty} \frac{1}{n}T_n = \mu\right\}$ then $P(\Omega_0) = 1$. From the definition of N_t, $T_{N_t} \le t < T_{N_t+1}$ and so

$$\frac{T_{N_t}}{N_t} \le \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t}.$$

Since $X_i > 0$ a.s., $\Omega_1 := \{N_t \uparrow \infty \text{ as } t \uparrow \infty\}$ also has full measure and for $\omega \in \Omega_0 \cap \Omega_1$ we have

$$\mu = \lim_{t\to\infty} \frac{T_{N_t(\omega)}(\omega)}{N_t(\omega)} \le \lim_{t\to\infty} \frac{t}{N_t(\omega)} \le \lim_{t\to\infty} \left[\frac{T_{N_t(\omega)+1}(\omega)}{N_t(\omega)+1} \frac{N_t(\omega)+1}{N_t(\omega)}\right] = \mu.$$

Example 12.20 (Renewal Theory II). Let $\{X_i\}_{i=1}^\infty$ be i.i.d. and $\{Y_i\}_{i=1}^\infty$ be i.i.d. with $\{X_i\}_{i=1}^\infty$ being independent of the $\{Y_i\}_{i=1}^\infty$. Also again assume that $0 < X_i < \infty$ and $0 < Y_i < \infty$ a.s. We will interpret Y_i to be the amount of time the i^{th} – bulb remains out after burning out before it is replaced by bulb number $i+1$. Let R_t be the amount of time that we have a working bulb in the time interval $[0,t]$. We are now going to show

$$\lim_{t\uparrow\infty} \frac{1}{t}R_t = \frac{\mathbb{E}X_1}{\mathbb{E}X_1 + \mathbb{E}Y_1}.$$

To prove this, now let $T_n := \sum_{i=1}^n (X_i + Y_i)$ be the time that the n^{th} – bulb is replaced and

$$N_t := \sup\{n \ge 0 : T_n \le t\}$$

denote the number of bulbs which have burned out up to time n. Then $R_t = \sum_{i=1}^{N_t} X_i$. Setting $\mu = \mathbb{E}X_1$ and $\nu = \mathbb{E}Y_1$, we now have $\frac{1}{t}N_t \to \frac{1}{\mu+\nu}$ a.s. so that $N_t = \frac{1}{\mu+\nu}t + o(t)$ a.s. Therefore, by the strong law of large numbers,

$$\frac{1}{t}R_t = \frac{1}{t}\sum_{i=1}^{N_t} X_i = \frac{N_t}{t} \cdot \frac{1}{N_t}\sum_{i=1}^{N_t} X_i \to \frac{1}{\mu+\nu} \cdot \mu \text{ a.s.}$$

Theorem 12.21 (Glivenko-Cantelli Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables and $F(x) := P(X_i \le x)$. Further let $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ be the* ***empirical distribution with empirical distribution function,***

$$F_n(x) := \mu_n((-\infty, x]) = \frac{1}{n}\sum_{i=1}^n 1_{X_i \le x}.$$

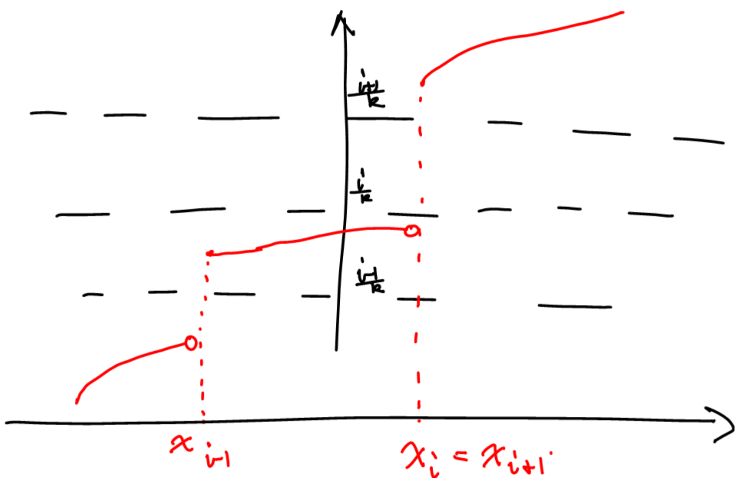
Then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \text{ a.s.}$$

Proof. Since $\{1_{X_i \le x}\}_{i=1}^{\infty}$ are i.i.d random variables with $\mathbb{E}1_{X_i \le x} = P(X_i \le x) = F(x)$, it follows by the strong law of large numbers the $\lim_{n\to\infty} F_n(x) = F(x)$ a.s. for each $x \in \mathbb{R}$. Our goal is to now show that this convergence is uniform.[1] To do this we will use one more application of the strong law of large numbers applied to $\{1_{X_i < x}\}$ which allows us to conclude, for each $x \in \mathbb{R}$, that

$$\lim_{n\to\infty} F_n(x-) = F(x-) \text{ a.s. (the null set depends on } x\text{).}$$

Given $k \in \mathbb{N}$, let $\Lambda_k := \left\{\frac{i}{k} : i = 1, 2, \ldots, k-1\right\}$ and let $x_i := \inf\{x : F(x) \ge i/k\}$ for $i = 1, 1, 2, \ldots, k-1$. Let us further set $x_k = \infty$ and $x_0 = -\infty$. Observe that it is possible that $x_i = x_{i+1}$ for some of the i. This can occur when F has jumps of size greater than $1/k$.



Now suppose i has been chosen so that $x_i < x_{i+1}$ and let $x \in (x_i, x_{i+1})$. Further let $N(\omega) \in \mathbb{N}$ be chosen so that

[1] Observation. If F is continouous then, by what we have just shown, there is a set $\Omega_0 \subset \Omega$ such that $P(\Omega_0) = 1$ and on Ω_0, $F_n(r) \to F(r)$ for all $r \in \mathbb{Q}$. Moreover on Ω_0, if $x \in \mathbb{R}$ and $r \le x \le s$ with $r, s \in \mathbb{Q}$, we have

$$F(r) = \lim_{n\to\infty} F_n(r) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le \lim_{n\to\infty} F_n(s) = F(s).$$

We may now let $s \downarrow x$ and $r \uparrow x$ to conclude, on Ω_0, on

$$F(x) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x) \text{ for all } x \in \mathbb{R},$$

i.e. on Ω_0, $\lim_{n\to\infty} F_n(x) = F(x)$. Thus, in this special case we have shown off a fixed null set independent of x that $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$.

$$|F_n(x_i) - F(x_i)| < 1/k \text{ and } |F_n(x_i-) - F(x_i-)| < 1/k$$

.

for $n \geq N(\omega)$ and $i = 1, 2, \ldots, k-1$ and $\omega \in \Omega_k$ with $P(\Omega_k) = 1$. We then have

$$F_n(x) \leq F_n(x_{i+1}-) \leq F(x_{i+1}-) + 1/k \leq F(x) + 2/k$$

and

$$F_n(x) \geq F_n(x_i) \geq F(x_i) - 1/k \geq F(x_{i+1}-) - 2/k \geq F(x) - 2/k.$$

From this it follows that $|F(x) - F_n(x)| \leq 2/k$ and we have shown for $\omega \in \Omega_k$ and $n \geq N(\omega)$ that

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \leq 2/k.$$

Hence it follows on $\Omega_0 := \cap_{k=1}^{\infty} \Omega_k$ (a set with $P(\Omega_0) = 1$) that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0.$$

■

Example 12.22 (Shannon's Theorem). Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with values in $\{1, 2, \ldots, r\} \subset \mathbb{N}$. Let $p(k) := P(X_i = k) > 0$ for $1 \leq k \leq r$. Further, let $\pi_n(\omega) = p(X_1(\omega)) \ldots p(X_n(\omega))$ be the probability of the realization, $(X_1(\omega), \ldots, X_n(\omega))$. Since $\{\ln p(X_i)\}_{i=1}^{\infty}$ are i.i.d.,

$$-\frac{1}{n} \ln \pi_n = -\frac{1}{n} \sum_{i=1}^{n} \ln p(X_i) \to -\mathbb{E}[\ln p(X_1)] = -\sum_{k=1}^{r} p(k) \ln p(k) =: H(p).$$

In particular if $\varepsilon > 0$, $P\left(\left|H - \frac{1}{n} \ln \pi_n\right| > \varepsilon\right) \to 0$ as $n \to \infty$. Since

$$\begin{aligned}\left\{\left|H + \frac{1}{n} \ln \pi_n\right| > \varepsilon\right\} &= \left\{H + \frac{1}{n} \ln \pi_n > \varepsilon\right\} \cup \left\{H + \frac{1}{n} \ln \pi_n < -\varepsilon\right\} \\ &= \left\{\frac{1}{n} \ln \pi_n > -H + \varepsilon\right\} \cup \left\{\frac{1}{n} \ln \pi_n < -H - \varepsilon\right\} \\ &= \left\{\pi_n > e^{n(-H+\varepsilon)}\right\} \cup \left\{\pi_n < e^{n(-H-\varepsilon)}\right\}\end{aligned}$$

and

$$\begin{aligned}\left\{\left|H - \frac{1}{n} \ln \pi_n\right| > \varepsilon\right\}^c &= \left\{\pi_n > e^{n(-H+\varepsilon)}\right\}^c \cup \left\{\pi_n < e^{n(-H-\varepsilon)}\right\}^c \\ &= \left\{\pi_n \leq e^{n(-H+\varepsilon)}\right\} \cap \left\{\pi_n \geq e^{n(-H-\varepsilon)}\right\} \\ &= \left\{e^{-n(H+\varepsilon)} \leq \pi_n \leq e^{-n(H-\varepsilon)}\right\},\end{aligned}$$

it follows that

$$P\left(e^{-n(H+\varepsilon)} \leq \pi_n \leq e^{-n(H-\varepsilon)}\right) \to 1 \text{ as } n \to \infty.$$

Thus the probability, π_n, that the random sample $\{X_1, \ldots, X_n\}$ should occur is approximately e^{-nH} with high probability. The number H is called the entropy of the distribution, $\{p(k)\}_{k=1}^r$.

12.4 More on the Weak Laws of Large Numbers

Theorem 12.23 (Weak Law of Large Numbers). *Suppose that* $\{X_n\}_{n=1}^\infty$ *is a sequence of independent random variables. Let* $S_n := \sum_{j=1}^n X_j$ *and*

$$a_n := \sum_{k=1}^n \mathbb{E}\left(X_k : |X_k| \leq n\right) = n\mathbb{E}\left(X_1 : |X_1| \leq n\right).$$

If

$$\sum_{k=1}^n P\left(|X_k| > n\right) \to 0 \tag{12.7}$$

and

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k| \leq n\right) \to 0, \tag{12.8}$$

then

$$\frac{S_n - a_n}{n} \stackrel{P}{\to} 0.$$

Proof. A key ingredient in this proof and proofs of other versions of the law of large numbers is to introduce truncations of the $\{X_k\}$. In this case we consider

$$S_n' := \sum_{k=1}^n X_k 1_{|X_k| \leq n}.$$

Since $\{S_n \neq S_{n'}\} \subset \cup_{k=1}^n \{|X_k| > n\}$,

$$P\left(\left|\frac{S_n - a_n}{n} - \frac{S_n' - a_n}{n}\right| > \varepsilon\right) = P\left(\left|\frac{S_n - S_n'}{n}\right| > \varepsilon\right)$$
$$\leq P\left(S_n \neq S_{n'}\right) \leq \sum_{k=1}^n P\left(|X_k| > n\right) \to 0 \text{ as } n \to \infty.$$

Hence it suffices to show $\frac{S_n' - a_n}{n} \stackrel{P}{\to} 0$ as $n \to \infty$ and for this it suffices to show, $\frac{S_n' - a_n}{n} \stackrel{L^2(P)}{\to} 0$ as $n \to \infty$.

Observe that $\mathbb{E}S_n' = a_n$ and therefore,

$$\mathbb{E}\left(\left[\frac{S_n' - a_n}{n}\right]^2\right) = \frac{1}{n^2}\operatorname{Var}(S_n') = \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}\left(X_k 1_{|X_k|\le n}\right)$$
$$\le \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 1_{|X_k|\le n}\right) \to 0 \text{ as } n\to\infty.$$

■

We now verify the hypothesis of Theorem 12.23 in three situations.

Corollary 12.24. *If $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^2(P)$ – random variables, then $\frac{1}{n}S_n \overset{P}{\to} \mu = \mathbb{E}X_1$.*

Proof. By the dominated convergence theorem,

$$\frac{a_n}{n} := \frac{1}{n}\sum_{k=1}^n \mathbb{E}(X_k : |X_k|\le n) = \mathbb{E}(X_1 : |X_1| \le n) \to \mu. \tag{12.9}$$

Moreover,

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k|\le n\right) = \frac{1}{n}\mathbb{E}\left(X_1^2 : |X_1|\le n\right) \le \frac{1}{n}\mathbb{E}\left(X_1^2\right) \to 0 \text{ as } n\to\infty$$

and by Chebyschev's inequality,

$$\sum_{k=1}^n P(|X_k| > n) = nP(|X_1|>n) \le n\frac{1}{n^2}\mathbb{E}|X_1|^2 \to 0 \text{ as } n\to\infty.$$

With these observations we may now apply Theorem 12.23 to complete the proof. ■

Corollary 12.25 (Khintchin's WLLN). *If $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^1(P)$ – random variables, then $\frac{1}{n}S_n \overset{P}{\to} \mu = \mathbb{E}X_1$.*

Proof. Again we have by Eq. (12.9), Chebyschev's inequality, and the dominated convergence theorem, that

$$\sum_{k=1}^n P(|X_k|>n) = nP(|X_1|>n) \le n\frac{1}{n}\mathbb{E}\left[|X_1| : |X_1|>n\right] \to 0 \text{ as } n\to\infty.$$

Also

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k|\le n\right) = \frac{1}{n}\mathbb{E}\left[|X_1|^2 : |X_1|\le n\right] = \mathbb{E}\left[|X_1|\frac{|X_1|}{n}1_{|X_1|\le n}\right]$$

and the latter expression goes to zero as $n\to\infty$ by the dominated convergence theorem, since

$$|X_1|\frac{|X_1|}{n}1_{|X_1|\le n} \le |X_1| \in L^1(P)$$

and $\lim_{n\to\infty} |X_1|\frac{|X_1|}{n}1_{|X_1|\le n} = 0$. Hence again the hypothesis of Theorem 12.23 have been verified. ■

Lemma 12.26. *Let X be a random variable such that $\tau(x) := xP(|X| \geq x) \to 0$ as $x \to \infty$, then*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[|X|^2 : |X| \leq n\right] = 0. \tag{12.10}$$

Note: *If $X \in L^1(P)$, then by Chebyschev's inequality and the dominated convergence theorem,*

$$\tau(x) \leq \mathbb{E}[|X| : |X| \geq x] \to 0 \text{ as } x \to \infty.$$

Proof. To prove this we observe that

$$\begin{aligned}\mathbb{E}\left[|X|^2 : |X| \leq n\right] &= \mathbb{E}\left[2\int 1_{0\leq x\leq |X|\leq n} x dx\right] = 2\int P(0 \leq x \leq |X| \leq n)\, x dx \\ &\leq 2\int_0^n xP(|X| \geq x)\, dx = 2\int_0^n \tau(x)\, dx.\end{aligned}$$

Now given $\varepsilon > 0$, let $M = M(\varepsilon)$ be chosen so that $\tau(x) \leq \varepsilon$ for $x \geq M$. Then

$$\mathbb{E}\left[|X|^2 : |X| \leq n\right] = 2\int_0^M \tau(x)\, dx + 2\int_M^n \tau(x)\, dx \leq 2KM + 2(n-M)\varepsilon$$

where $K = \sup\{\tau(x) : x \geq 0\}$. Dividing this estimate by n and then letting $n \to \infty$ shows

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\left[|X|^2 : |X| \leq n\right] \leq 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. ■

Corollary 12.27 (Feller's WLLN). *If $\{X_n\}_{n=1}^\infty$ are i.i.d. and $\tau(x) := xP(|X_1| > x) \to 0$ as $x \to \infty$, then the hypothesis of Theorem 12.23 are satisfied.*

Proof. Since

$$\sum_{k=1}^n P(|X_k| > n) = nP(|X_1| > n) = \tau(n) \to 0 \text{ as } n \to \infty,$$

Eq. (12.7) is satisfied. Eq. (12.8), follows from Lemma 12.26 and the identity,

$$\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 : |X_k| \leq n\right) = \frac{1}{n}\mathbb{E}\left[|X_1|^2 : |X_1| \leq n\right].$$

■

12.5 Maximal Inequalities

Theorem 12.28 (Kolmogorov's Inequality). *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, $S_n := X_1 + \cdots + X_n$, and $S_n^* = \max_{j\le n}|S_j|$. Then for any $\alpha > 0$ we have*

$$P(S_N^* \ge \alpha) \le \frac{1}{\alpha^2}\mathbb{E}\left[S_N^2 : |S_N^*| \ge \alpha\right].$$

(See Proposition 19.38 and Example 19.40 below for generalizations of this inequality.)

Proof. Let $J = \inf\{j : |S_j| \ge \alpha\}$ with the infimum of the empty set being taken to be equal to ∞. Observe that

$$\{J = j\} = \{|S_1| < \alpha, \ldots, |S_{j-1}| < \alpha, |S_j| \ge \alpha\} \in \sigma(X_1, \ldots, X_j).$$

Now

$$\begin{aligned}
\mathbb{E}\left[S_N^2 : |S_N^*| > \alpha\right] &= \mathbb{E}\left[S_N^2 : J \le N\right] = \sum_{j=1}^N \mathbb{E}\left[S_N^2 : J = j\right] \\
&= \sum_{j=1}^N \mathbb{E}\left[(S_j + S_N - S_j)^2 : J = j\right] \\
&= \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 + 2S_j(S_N - S_j) : J = j\right] \\
&\overset{(*)}{=} \sum_{j=1}^N \mathbb{E}\left[S_j^2 + (S_N - S_j)^2 : J = j\right] \\
&\ge \sum_{j=1}^N \mathbb{E}\left[S_j^2 : J = j\right] \ge \alpha^2 \sum_{j=1}^N P[J = j] = \alpha^2 P(|S_N^*| > \alpha).
\end{aligned}$$

The equality, $(*)$, is a consequence of the observations: 1) $1_{J=j}S_j$ is $\sigma(X_1, \ldots, X_j)$ – measurable, 2) $(S_n - S_j)$ is $\sigma(X_{j+1}, \ldots, X_n)$ – measurable and hence $1_{J=j}S_j$ and $(S_n - S_j)$ are independent, and so 3)

$$\begin{aligned}
\mathbb{E}[S_j(S_N - S_j) : J = j] &= \mathbb{E}[S_j 1_{J=j}(S_N - S_j)] \\
&= \mathbb{E}[S_j 1_{J=j}] \cdot \mathbb{E}[S_N - S_j] = \mathbb{E}[S_j 1_{J=j}] \cdot 0 = 0.
\end{aligned}$$

■

Corollary 12.29 (L^2 – SSLN). *Let $\{X_n\}$ be a sequence of independent random variables with mean zero, and $\sigma^2 = \mathbb{E}X_n^2 < \infty$. Letting $S_n = \sum_{k=1}^n X_k$ and $p > 1/2$, we have*

$$\frac{1}{n^p} S_n \to 0 \ a.s.$$

If $\{Y_n\}$ *is a sequence of independent random variables* $\mathbb{E}Y_n = \mu$ *and* $\sigma^2 = \operatorname{Var}(X_n) < \infty$, *then for any* $\beta \in (0, 1/2)$,

$$\frac{1}{n} \sum_{k=1}^{n} Y_k - \mu = O\left(\frac{1}{n^\beta}\right).$$

Proof. (The proof of this Corollary may be skipped. We will give another proof in Corollary 12.36 below.) From Theorem 12.28, we have for every $\varepsilon > 0$ that

$$P\left(\frac{S_N^*}{N^p} \geq \varepsilon\right) = P\left(S_N^* \geq \varepsilon N^p\right) \leq \frac{1}{\varepsilon^2 N^{2p}} \mathbb{E}\left[S_N^2\right] = \frac{1}{\varepsilon^2 N^{2p}} CN = \frac{C}{\varepsilon^2 N^{(2p-1)}}.$$

Hence if we suppose that $N_n = n^\alpha$ with $\alpha(2p-1) > 1$, then we have

$$\sum_{n=1}^{\infty} P\left(\frac{S_{N_n}^*}{N_n^p} \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{C}{\varepsilon^2 n^{\alpha(2p-1)}} < \infty$$

and so by the first Borel – Cantelli lemma we have

$$P\left(\left\{\frac{S_{N_n}^*}{N_n^p} \geq \varepsilon \text{ for } n \text{ i.o.}\right\}\right) = 0.$$

From this it follows that $\lim_{n\to\infty} \frac{S_{N_n}^*}{N_n^p} = 0$ a.s.

To finish the proof, for $m \in \mathbb{N}$, we may choose $n = n(m)$ such that

$$n^\alpha = N_n \leq m < N_{n+1} = (n+1)^\alpha.$$

Since

$$\frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \leq \frac{S_m^*}{m^p} \leq \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p}$$

and

$$N_{n+1}/N_n \to 1 \text{ as } n \to \infty,$$

it follows that

$$\begin{aligned} 0 = \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)}^p} &= \lim_{m\to\infty} \frac{S_{N_{n(m)}}^*}{N_{n(m)+1}^p} \leq \lim_{m\to\infty} \frac{S_m^*}{m^p} \\ &\leq \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)}^p} = \lim_{m\to\infty} \frac{S_{N_{n(m)+1}}^*}{N_{n(m)+1}^p} = 0 \text{ a.s.} \end{aligned}$$

That is $\lim_{m\to\infty} \frac{S_m^*}{m^p} = 0$ a.s. ∎

Theorem 12.30 (Skorohod's Inequality). *Let $\{X_n\}$ be a sequence of independent random variables and let $\alpha > 0$. Let $S_n := X_1 + \cdots + X_n$. Then for all $\alpha > 0$,*

$$P\left(|S_N| > \alpha\right) \geq \left(1 - c_N\left(\alpha\right)\right) P\left(\max_{j \leq N} |S_j| > 2\alpha\right),$$

where

$$c_N \ \left(\alpha\right) := \max_{j \leq N} P\left(|S_N - S_j| > \alpha\right).$$

Proof. Our goal is to compute

$$P\left(\max_{j \leq N} |S_j| > 2\alpha\right).$$

To this end, let $J = \inf\{j : |S_j| > 2\alpha\}$ with the infimum of the empty set being taken to be equal to ∞. Observe that

$$\{J = j\} = \{|S_1| \leq 2\alpha, \ldots, |S_{j-1}| \leq 2\alpha, |S_j| > 2\alpha\}$$

and therefore

$$\left\{\max_{j \leq N} |S_j| > 2\alpha\right\} = \sum_{j=1}^{N} \{J = j\}.$$

Also observe that on $\{J = j\}$,

$$|S_N| = |S_N - S_j + S_j| \geq |S_j| - |S_N - S_j| > 2\alpha - |S_N - S_j|.$$

Hence on the $\{J = j, |S_N - S_j| \leq \alpha\}$ we have $|S_N| > \alpha$, i.e.

$$\{J = j, |S_N - S_j| \leq \alpha\} \subset \{|S_N| > \alpha\} \text{ for all } j \leq N.$$

Hence ti follows from this identity and the independence of $\{X_n\}$ that

$$\begin{aligned} P\left(|S_N| > \alpha\right) &\geq \sum_{j=1}^{N} P\left(J = j, |S_N - S_j| \leq \alpha\right) \\ &= \sum_{j=1}^{N} P\left(J = j\right) P\left(|S_N - S_j| \leq \alpha\right). \end{aligned}$$

Under the assumption that $P\left(|S_N - S_j| > \alpha\right) \leq c$ for all $j \leq N$, we find

$$P\left(|S_N - S_j| \leq \alpha\right) \geq 1 - c$$

and therefore,

$$P\left(|S_N| > \alpha\right) \geq \sum_{j=1}^{N} P\left(J = j\right)(1 - c) = (1 - c) P\left(\max_{j \leq N} |S_j| > 2\alpha\right).$$

■

As an application of Theorem 12.30 we have the following convergence result.

Theorem 12.31 (Lévy's Theorem). *Suppose that $\{X_n\}_{n=1}^\infty$ are i.i.d. random variables then $\sum_{n=1}^\infty X_n$ converges in probability iff $\sum_{n=1}^\infty X_n$ converges a.s.*

Proof. Let $S_n := \sum_{k=1}^n X_k$. Since almost sure convergence implies convergence in probability, it suffices to show; if S_n is convergent in probability then S_n is almost surely convergent. Given $M \in \mathbb{M}$, let $Q_M := \sup_{n\geq M} |S_n - S_M|$ and for $M < N$, let $Q_{M,N} := \sup_{M\leq n\leq N} |S_n - S_M|$. Given $\varepsilon \in (0,1)$, by assumption, there exists $M = M(\varepsilon) \in \mathbb{N}$ such that $\max_{M\leq j\leq N} P(|S_N - S_j| > \varepsilon) < \varepsilon$ for all $N \geq M$. An application of Skorohod's inequality, then shows

$$P(Q_{M,N} \geq 2\varepsilon) \leq \frac{P(|S_N - S_M| > \varepsilon)}{(1 - \max_{M\leq j\leq N} P(|S_N - S_j| > \varepsilon))} \leq \frac{\varepsilon}{1-\varepsilon}.$$

Since $Q_{M,N} \uparrow Q_M$ as $N \to \infty$, we may conclude

$$P(Q_M \geq 2\varepsilon) \leq \frac{\varepsilon}{1-\varepsilon}.$$

Since,

$$\delta_M := \sup_{m,n\geq M} |S_n - S_m| \leq \sup_{m,n\geq M} [|S_n - S_M| + |S_M - S_m|] = 2Q_M$$

we may further conclude, $P(\delta_M > 4\varepsilon) \leq \frac{\varepsilon}{1-\varepsilon}$ and since $\varepsilon > 0$ is arbitrary, it follows that $\delta_M \xrightarrow{P} 0$ as $M \to \infty$. Moreover, since δ_M is decreasing in M, it follows that $\lim_{M\to\infty} \delta_M =: \delta$ exists and because $\delta_M \xrightarrow{P} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n\to\infty} |S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^\infty$ is almost surely Cauchy and hence almost surely convergent. ■

Proposition 12.32 (Reflection Principle). *Let X be a separable Banach space and $\{\xi_i\}_{i=1}^N$ be independent symmetric (i.e. $\xi_i \stackrel{d}{=} -\xi_i$) random variables with values in X. Let $S_k := \sum_{i=1}^k \xi_i$ and $S_k^* := \sup_{j\leq k} \|S_j\|$ with the convention that $S_0^* = 0$. Then*

$$P(S_N^* \geq r) \leq 2P(\|S_N\| \geq r). \tag{12.11}$$

Proof. Since

$$\{S_N^* \geq r\} = \sum\nolimits_{j=1}^N \left\{\|S_j\| \geq r,\ S_{j-1}^* < r\right\},$$

$$\begin{aligned} P(S_N^* \geq r) &= P(S_N^* \geq r, \ \|S_N\| \geq r) + P(S_N^* \geq r, \ \|S_N\| < r) \\ &= P(\|S_N\| \geq r) + P(S_N^* \geq r, \ \|S_N\| < r). \end{aligned} \tag{12.12}$$

where

$$P(S_N^* \geq r, \ \|S_N\| < r) = \sum_{j=1}^{N} P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| < r). \tag{12.13}$$

By symmetry and independence we have

$$\begin{aligned} P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| < r) &= P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \left\|S_j + \sum_{k>j} \xi_k\right\| < r) \\ &= P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \left\|S_j - \sum_{k>j} \xi_k\right\| < r) \\ &= P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|2S_j - S_N\| < r). \end{aligned}$$

If $\|S_j\| \geq r$ and $\|2S_j - S_N\| < r$, then

$$r > \|2S_j - S_N\| \geq 2\|S_j\| - \|S_N\| \geq 2r - \|S_N\|$$

and hence $\|S_N\| > r$. This shows,

$$\left\{\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|2S_j - S_N\| < r\right\} \subset \left\{\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| > r\right\}$$

and therefore,

$$P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| < r) \leq P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| > r).$$

Combining the estimate with Eq. (12.13) gives

$$\begin{aligned} P(S_N^* \geq r, \ \|S_N\| < r) &\leq \sum_{j=1}^{N} P(\|S_j\| \geq r, \ S_{j-1}^* < r, \ \|S_N\| > r) \\ &= P(S_N^* \geq r, \ \|S_N\| > r) \leq P(\|S_N\| \geq r). \end{aligned}$$

This estimate along with the estimate in Eq. (12.12) completes the proof of the theorem. ■

12.6 Kolmogorov's Convergence Criteria and the SSLN

We are now in a position to prove Theorem 12.11 which we restate here.

Theorem 12.33 (Kolmogorov's Convergence Criteria). *Suppose that* $\{Y_n\}_{n=1}^{\infty}$ *are independent square integrable random variables. If* $\sum_{j=1}^{\infty} \operatorname{Var}(Y_j) < \infty$, *then* $\sum_{j=1}^{\infty} (Y_j - \mathbb{E}Y_j)$ *converges a.s.*

Proof. First proof. By Proposition 12.5, the sum, $\sum_{j=1}^{\infty}(Y_j - \mathbb{E}Y_j)$, is $L^2(P)$ convergent and hence convergent in probability. An application of Lévy's Theorem 12.31 then shows $\sum_{j=1}^{\infty}(Y_j - \mathbb{E}Y_j)$ is almost surely convergent.

Second proof. Let $S_n := \sum_{j=1}^{n} X_j$ where $X_j := Y_j - \mathbb{E}Y_j$. According to Kolmogorov's inequality, Theorem 12.28, for all $M < N$,

$$P\left(\max_{M\le j\le N}|S_j - S_M| \ge \alpha\right) \le \frac{1}{\alpha^2}\mathbb{E}\left[(S_N - S_M)^2\right] = \frac{1}{\alpha^2}\sum_{j=M+1}^{N}\mathbb{E}\left[X_j^2\right]$$
$$= \frac{1}{\alpha^2}\sum_{j=M+1}^{N}\mathrm{Var}(X_j).$$

Letting $N \to \infty$ in this inequality shows, with $Q_M := \sup_{j\ge M}|S_j - S_M|$,

$$P(Q_M \ge \alpha) \le \frac{1}{\alpha^2}\sum_{j=M+1}^{\infty}\mathrm{Var}(X_j).$$

Since

$$\delta_M := \sup_{j,k\ge M}|S_j - S_k| \le \sup_{j,k\ge M}\left[|S_j - S_M| + |S_M - S_k|\right] \le 2Q_M$$

we may further conclude,

$$P(\delta_M \ge 2\alpha) \le \frac{1}{\alpha^2}\sum_{j=M+1}^{\infty}\mathrm{Var}(X_j) \to 0 \text{ as } M \to \infty,$$

i.e. $\delta_M \xrightarrow{P} 0$ as $M \to \infty$. Since δ_M is decreasing in M, it follows that $\lim_{M\to\infty}\delta_M =: \delta$ exists and because $\delta_M \xrightarrow{P} 0$ we may concluded that $\delta = 0$ a.s. Thus we have shown

$$\lim_{m,n\to\infty}|S_n - S_m| = 0 \text{ a.s.}$$

and therefore $\{S_n\}_{n=1}^{\infty}$ is almost surely Cauchy and hence almost surely convergent. ∎

Lemma 12.34 (Kronecker's Lemma). *Suppose that* $\{x_k\} \subset \mathbb{R}$ *and* $\{a_k\} \subset (0,\infty)$ *are sequences such that* $a_k \uparrow \infty$ *and* $\sum_{k=1}^{\infty}\frac{x_k}{a_k}$ *exists. Then*

$$\lim_{n\to\infty}\frac{1}{a_n}\sum_{k=1}^{n}x_k = 0.$$

Proof. Before going to the proof, let us warm-up by proving the following continuous version of the lemma. Let $a(s) \in (0, \infty)$ and $x(s) \in \mathbb{R}$ be continuous functions such that $a(s) \uparrow \infty$ as $s \to \infty$ and $\int_1^\infty \frac{x(s)}{a(s)} ds$ exists. We are going to show

$$\lim_{n\to\infty} \frac{1}{a(n)} \int_1^n x(s)\, ds = 0.$$

Let $X(s) := \int_0^s x(u)\, du$ and

$$r(s) := \int_s^\infty \frac{X'(u)}{a(u)} du = \int_s^\infty \frac{x(u)}{a(u)} du.$$

Then by assumption, $r(s) \to 0$ as $s \to 0$ and $X'(s) = -a(s) r'(s)$. Integrating this equation shows

$$X(s) - X(s_0) = -\int_{s_0}^s a(u) r'(u)\, du = -a(u) r(u) |_{u=s_0}^s + \int_{s_0}^s r(u) a'(u)\, du.$$

Dividing this equation by $a(s)$ and then letting $s \to \infty$ gives

$$\begin{aligned}
\limsup_{s\to\infty} \frac{|X(s)|}{a(s)} &= \limsup_{s\to\infty} \left[\frac{a(s_0) r(s_0) - a(s) r(s)}{a(s)} + \frac{1}{a(s)} \int_{s_0}^s r(u) a'(u)\, du \right] \\
&\leq \limsup_{s\to\infty} \left[-r(s) + \frac{1}{a(s)} \int_{s_0}^s |r(u)| a'(u)\, du \right] \\
&\leq \limsup_{s\to\infty} \left[\frac{a(s) - a(s_0)}{a(s)} \sup_{u \geq s_0} |r(u)| \right] = \sup_{u\geq s_0} |r(u)| \to 0 \text{ as } s_0 \to \infty.
\end{aligned}$$

With this as warm-up, we go to the discrete case.

Let

$$S_k := \sum_{j=1}^k x_j \text{ and } r_k := \sum_{j=k}^\infty \frac{x_j}{a_j}.$$

so that $r_k \to 0$ as $k \to \infty$ by assumption. Since $x_k = a_k (r_k - r_{k+1})$, we find

$$\begin{aligned}
\frac{S_n}{a_n} &= \frac{1}{a_n} \sum_{k=1}^n a_k (r_k - r_{k+1}) = \frac{1}{a_n} \left[\sum_{k=1}^n a_k r_k - \sum_{k=2}^{n+1} a_{k-1} r_k \right] \\
&= \frac{1}{a_n} \left[a_1 r_1 - a_n r_{n+1} + \sum_{k=2}^n (a_k - a_{k-1}) r_k \right]. \text{ (summation by parts)}
\end{aligned}$$

Using the fact that $a_k - a_{k-1} \geq 0$ for all $k \geq 2$, and

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=2}^m (a_k - a_{k-1}) |r_k| = 0$$

for any $m \in \mathbb{N}$; we may conclude

$$\begin{aligned}
\limsup_{n\to\infty}\left|\frac{S_n}{a_n}\right| &\le \limsup_{n\to\infty}\frac{1}{a_n}\left[\sum_{k=2}^{n}(a_k-a_{k-1})\,|r_k|\right] \\
&= \limsup_{n\to\infty}\frac{1}{a_n}\left[\sum_{k=m}^{n}(a_k-a_{k-1})\,|r_k|\right] \\
&\le \sup_{k\ge m}|r_k|\cdot\limsup_{n\to\infty}\frac{1}{a_n}\left[\sum_{k=m}^{n}(a_k-a_{k-1})\right] \\
&= \sup_{k\ge m}|r_k|\cdot\limsup_{n\to\infty}\frac{1}{a_n}\left[a_n-a_{m-1}\right] = \sup_{k\ge m}|r_k|\,.
\end{aligned}$$

This completes the proof since $\sup_{k\ge m}|r_k|\to 0$ as $m\to\infty$. ∎

Corollary 12.35. *Let $\{X_n\}$ be a sequence of independent square integrable random variables and b_n be a sequence such that $b_n\uparrow\infty$. If*

$$\sum_{k=1}^{\infty}\frac{\operatorname{Var}(X_k)}{b_k^2}<\infty$$

then

$$\frac{S_n-\mathbb{E}S_n}{b_n}\to 0 \text{ a.s.}$$

Proof. By Kolmogorov's Convergence Criteria, Theorem 12.33,

$$\sum_{k=1}^{\infty}\frac{X_k-\mathbb{E}X_k}{b_k}\text{ is convergent a.s.}$$

Therefore an application of Kronecker's Lemma implies

$$0=\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=1}^{n}(X_k-\mathbb{E}X_k)=\lim_{n\to\infty}\frac{S_n-\mathbb{E}S_n}{b_n}.$$

∎

Corollary 12.36 (L^2 – SSLN). *Let $\{X_n\}$ be a sequence of independent random variables such that $\sigma^2=\mathbb{E}X_n^2<\infty$. Letting $S_n=\sum_{k=1}^{n}X_k$ and $\mu:=\mathbb{E}X_n$, we have*

$$\frac{1}{b_n}(S_n-n\mu)\to 0 \text{ a.s.} \tag{12.14}$$

provided $b_n\uparrow\infty$ and $\sum_{n=1}^{\infty}\frac{1}{b_n^2}<\infty$. For example, we could take $b_n=n$ or $b_n=n^p$ for an $p>1/2$, or $b_n=n^{1/2}(\ln n)^{1/2+\varepsilon}$ for any $\varepsilon>0$. We may rewrite Eq. (12.14) as

$$S_n-n\mu=o(1)\,b_n$$

or equivalently,

$$\frac{S_n}{n}-\mu=o(1)\,\frac{b_n}{n}.$$

Proof. This corollary is a special case of Corollary 12.35. Let us simply observe here that

$$\sum_{n=2}^{\infty} \frac{1}{\left(n^{1/2} (\ln n)^{1/2+\varepsilon}\right)^2} = \sum_{n=2}^{\infty} \frac{1}{n (\ln n)^{1+2\varepsilon}}$$

by comparison with the integral

$$\int_2^\infty \frac{1}{x \ln^{1+2\varepsilon} x} dx = \int_{\ln 2}^\infty \frac{1}{e^y y^{1+2\varepsilon}} e^y dy = \int_{\ln 2}^\infty \frac{1}{y^{1+2\varepsilon}} dy < \infty,$$

wherein we have made the change of variables, $y = \ln x$. ■

Fact 12.37 *Under the hypothesis in Corollary 12.36,*

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n^{1/2} (\ln \ln n)^{1/2}} = \sqrt{2}\sigma \ \ a.s.$$

Our next goal is to prove the Strong Law of Large numbers (in Theorem 12.7) under the assumption that $\mathbb{E}|X_1| < \infty$.

12.7 Strong Law of Large Numbers

Lemma 12.38. *Suppose that $X : \Omega \to \mathbb{R}$ is a random variable, then*

$$\mathbb{E}|X|^p = \int_0^\infty ps^{p-1} P(|X| \geq s)\, ds = \int_0^\infty ps^{p-1} P(|X| > s)\, ds.$$

Proof. By the fundamental theorem of calculus,

$$|X|^p = \int_0^{|X|} ps^{p-1} ds = p\int_0^\infty 1_{s\leq|X|} \cdot s^{p-1} ds = p\int_0^\infty 1_{s<|X|} \cdot s^{p-1} ds.$$

Taking expectations of this identity along with an application of Tonelli's theorem completes the proof. ■

Lemma 12.39. *If X is a random variable and $\varepsilon > 0$, then*

$$\sum_{n=1}^{\infty} P(|X| \geq n\varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}|X| \leq \sum_{n=0}^{\infty} P(|X| \geq n\varepsilon). \tag{12.15}$$

Proof. First observe that for all $y \geq 0$ we have,

$$\sum_{n=1}^{\infty} 1_{n\leq y} \leq y \leq \sum_{n=1}^{\infty} 1_{n\leq y} + 1 = \sum_{n=0}^{\infty} 1_{n\leq y}. \tag{12.16}$$

Taking $y = |X|/\varepsilon$ in Eq. (12.16) and then take expectations gives the estimate in Eq. (12.15). ■

Proposition 12.40. *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables, then the following are equivalent:*

1. $\mathbb{E}|X_1| < \infty$.
2. *There exists* $\varepsilon > 0$ *such that* $\sum_{n=1}^{\infty} P(|X_1| \geq \varepsilon n) < \infty$.
3. *For all* $\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|X_1| \geq \varepsilon n) < \infty$.
4. $\lim_{n\to\infty} \frac{|X_n|}{n} = 0$ *a.s.*

Proof. The equivalence of items 1., 2., and 3. easily follows from Lemma 12.39. So to finish the proof it suffices to show 3. is equivalent to 4. To this end we start by noting that $\lim_{n\to\infty} \frac{|X_n|}{n} = 0$ a.s. iff

$$0 = P\left(\frac{|X_n|}{n} \geq \varepsilon \text{ i.o.}\right) = P(|X_n| \geq n\varepsilon \text{ i.o.}) \text{ for all } \varepsilon > 0. \tag{12.17}$$

However, since $\{|X_n| \geq n\varepsilon\}_{n=1}^{\infty}$ are independent sets, Borel zero-one law shows the statement in Eq. (12.17) is equivalent to $\sum_{n=1}^{\infty} P(|X_n| \geq n\varepsilon) < \infty$ for all $\varepsilon > 0$. ∎

Corollary 12.41. *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables such that $\frac{1}{n}S_n \to c \in \mathbb{R}$ a.s., then $X_n \in L^1(P)$ and $\mu := \mathbb{E}X_n = c$.*

Proof. If $\frac{1}{n}S_n \to c$ a.s. then $\varepsilon_n := \frac{S_{n+1}}{n+1} - \frac{S_n}{n} \to 0$ a.s. and therefore,

$$\begin{aligned}\frac{X_{n+1}}{n+1} &= \frac{S_{n+1}}{n+1} - \frac{S_n}{n+1} = \varepsilon_n + S_n\left[\frac{1}{n} - \frac{1}{n+1}\right] \\ &= \varepsilon_n + \frac{1}{(n+1)}\frac{S_n}{n} \to 0 + 0 \cdot c = 0.\end{aligned}$$

Hence an application of Proposition 12.40 shows $X_n \in L^1(P)$. Moreover by Exercise 11.3, $\left\{\frac{1}{n}S_n\right\}_{n=1}^{\infty}$ is a uniformly integrable sequenced and therefore,

$$\mu = \mathbb{E}\left[\frac{1}{n}S_n\right] \to \mathbb{E}\left[\lim_{n\to\infty}\frac{1}{n}S_n\right] = \mathbb{E}[c] = c.$$

∎

Lemma 12.42. *For all $x \geq 0$,*

$$\varphi(x) := \sum_{n=1}^{\infty} \frac{1}{n^2} 1_{x \leq n} = \sum_{n \geq x} \frac{1}{n^2} \leq 2 \cdot \min\left(\frac{1}{x}, 1\right).$$

Proof. The proof will be by comparison with the integral, $\int_a^{\infty} \frac{1}{t^2} dt = 1/a$. For example,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \leq 1 + \int_1^{\infty} \frac{1}{t^2} dt = 1 + 1 = 2$$

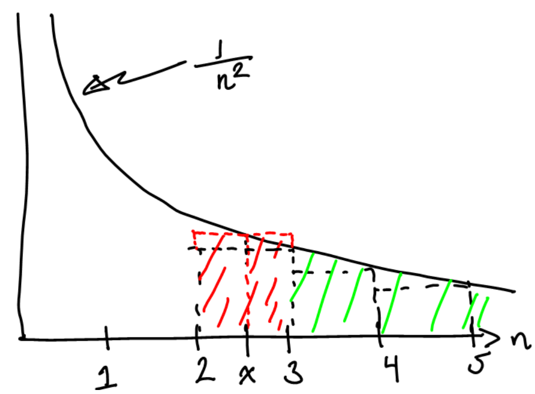
and so

$$\sum_{n \geq x} \frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} = 2 \leq \frac{2}{x} \text{ for } 0 < x \leq 1.$$

Similarly, for $x > 1$,

$$\sum_{n \geq x} \frac{1}{n^2} \leq \frac{1}{x^2} + \int_x^{\infty} \frac{1}{t^2} dt = \frac{1}{x^2} + \frac{1}{x} = \frac{1}{x}\left(1 + \frac{1}{x}\right) \leq \frac{2}{x},$$

see Figure 12.7 below. ■



Lemma 12.43. *Suppose that $X : \Omega \to \mathbb{R}$ is a random variable, then*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\left[|X|^2 : 1_{|X| \leq n}\right] \leq 2\mathbb{E}\,|X|\,.$$

Proof. This is a simple application of Lemma 12.42;

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{E}\left[|X|^2 : 1_{|X| \leq n}\right] &= \mathbb{E}\left[|X|^2 \sum_{n=1}^{\infty} \frac{1}{n^2} 1_{|X| \leq n}\right] = \mathbb{E}\left[|X|^2 \varphi\left(|X|\right)\right] \\ &\leq 2\mathbb{E}\left[|X|^2 \left(\frac{1}{|X|} \wedge 1\right)\right] \leq 2\mathbb{E}\,|X|\,.\end{aligned}$$

■

With this as preparation we are now in a position to prove Theorem 12.7 which we restate here.

Theorem 12.44 (Kolmogorov's Strong Law of Large Numbers). *Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables and let $S_n := X_1 + \cdots + X_n$. Then there exists $\mu \in \mathbb{R}$ such that $\frac{1}{n} S_n \to \mu$ a.s. iff X_n is integrable and in which case $\mathbb{E}X_n = \mu$.*

Proof. The implication, $\frac{1}{n}S_n \to \mu$ a.s. implies $X_n \in L^1(P)$ and $\mathbb{E}X_n = \mu$ has already been proved in Corollary 12.41. So let us now assume $X_n \in L^1(P)$ and let $\mu := \mathbb{E}X_n$.

Let $X_n' := X_n 1_{|X_n| \le n}$. By Proposition 12.40,

$$\sum_{n=1}^{\infty} P(X_n' \neq X_n) = \sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X_1| > n) \le \mathbb{E}|X_1| < \infty,$$

and hence $\{X_n\}$ and $\{X_n'\}$ are tail equivalent. Therefore it suffices to show $\lim_{n\to\infty} \frac{1}{n} S_n' = \mu$ a.s. where $S_n' := X_1' + \cdots + X_n'$. But by Lemma 12.43,

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n')}{n^2} \le \sum_{n=1}^{\infty} \frac{\mathbb{E}|X_n'|^2}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[|X_n|^2 1_{|X_n| \le n}\right]}{n^2}$$
$$= \sum_{n=1}^{\infty} \frac{\mathbb{E}\left[|X_1|^2 1_{|X_1| \le n}\right]}{n^2} \le 2\mathbb{E}|X_1| < \infty.$$

Therefore by Kolmogorov's convergence criteria,

$$\sum_{n=1}^{\infty} \frac{X_n' - \mathbb{E}X_n'}{n} \text{ is almost surely convergent.}$$

Kronecker's lemma then implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (X_k' - \mathbb{E}X_k') = 0 \text{ a.s.}$$

So to finish the proof, it only remains to observe

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}X_k' = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_n 1_{|X_n| \le n}\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[X_1 1_{|X_1| \le n}\right]$$
$$= \lim_{n\to\infty} \mathbb{E}\left[X_1 1_{|X_1| \le n}\right] = \mu.$$

Here we have used the dominated convergence theorem to see that $a_n := \mathbb{E}\left[X_1 1_{|X_1| \le n}\right] \to \mu$ as $n \to \infty$. It is now easy (and standard) to check that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_n = \lim_{n\to\infty} a_n = \mu$ as well. ■

We end this section with another example of using Kolmogorov's convergence criteria in conjunction with Kronecker's lemma. We now assume that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random variables with a continuous distribution function and let A_j denote the event when X_j is a record, i.e.

$$A_j := \{X_j > \max\{X_1, X_2, \ldots, X_{k-1}\}\}.$$

Recall from Renyi Theorem 7.28 that $\{A_j\}_{j=1}^{\infty}$ are independent and $P(A_j) = \frac{1}{j}$ for all j.

Proposition 12.45. *Keeping the preceding notation and let* $\mu_N := \sum_{j=1}^N 1_{A_j}$ *denote the number of records in the first* N *observations. Then* $\lim_{N\to\infty} \frac{\mu_N}{\ln N} = 1$ *a.s.*

Proof. Since 1_{A_j} are Bernoulli random variables, $\mathbb{E}1_{A_j} = \frac{1}{j}$ and

$$\operatorname{Var}\left(1_{A_j}\right) = \mathbb{E}1_{A_j}^2 - \left(\mathbb{E}1_{A_j}\right)^2 = \frac{1}{j} - \frac{1}{j^2} = \frac{j-1}{j^2}.$$

Observing that

$$\sum_{j=1}^n \mathbb{E}1_{A_j} = \sum_{j=1}^n \frac{1}{j} \sim \int_1^N \frac{1}{x} dx = \ln N$$

we are lead to try to normalize the sum $\sum_{j=1}^N 1_{A_j}$ by $\ln N$. So in the spirit of the proof of the strong law of large numbers let us compute;

$$\sum_{j=2}^\infty \operatorname{Var}\left(\frac{1_{A_j}}{\ln j}\right) = \sum_{j=2}^\infty \frac{1}{\ln^2 j} \frac{j-1}{j^2} \sim \int_2^\infty \frac{1}{\ln^2 x} \frac{1}{x} dx = \int_{\ln 2}^\infty \frac{1}{y^2} dy < \infty.$$

Therefore by Kolmogorov's convergence criteria we may conclude

$$\sum_{j=2}^\infty \frac{1_{A_j} - \frac{1}{j}}{\ln j} = \sum_{j=2}^\infty \left[\frac{1_{A_j}}{\ln j} - \mathbb{E}\left[\frac{1_{A_j}}{\ln j}\right]\right]$$

is almost surely convergent. An application of Kronecker's Lemma then implies

$$\lim_{n\to\infty} \frac{\sum_{j=1}^N \left(1_{A_j} - \frac{1}{j}\right)}{\ln N} = 0 \text{ a.s.}$$

So to finish the proof it only remains to show

$$\lim_{n\to\infty} \frac{\sum_{j=1}^N \frac{1}{j}}{\ln N} = 1. \tag{12.18}$$

To see this write

$$\begin{aligned} \ln(N+1) &= \int_1^{N+1} \frac{1}{x} dx = \sum_{j=1}^N \int_j^{j+1} \frac{1}{x} dx \\ &= \sum_{j=1}^N \int_j^{j+1} \left(\frac{1}{x} - \frac{1}{j}\right) dx + \sum_{j=1}^N \frac{1}{j} \\ &= \rho_N + \sum_{j=1}^N \frac{1}{j} \end{aligned} \tag{12.19}$$

where

$$|\rho_N| = \sum_{j=1}^{N} \left| \ln \frac{j+1}{j} - \frac{1}{j} \right| = \sum_{j=1}^{N} \left| \ln (1 + 1/j) - \frac{1}{j} \right| \sim \sum_{j=1}^{N} \frac{1}{j^2}$$

and hence we conclude that $\lim_{N\to\infty} \rho_N < \infty$. So dividing Eq. (12.19) by $\ln N$ and letting $N \to \infty$ gives the desired limit in Eq. (12.18). ∎

12.8 Necessity Proof of Kolmogorov's Three Series Theorem

This section is devoted to the necessity part of the proof of Kolmogorov's Three Series Theorem 12.12. We start with a couple of lemmas.

Lemma 12.46. *Suppose that $\{Y_n\}_{n=1}^{\infty}$ are independent random variables such that there exists $c < \infty$ such that $|Y_n| \le c < \infty$ a.s. and further assume $\mathbb{E}Y_n = 0$. If $\sum_{n=1}^{\infty} Y_n$ is almost surely convergent then $\sum_{n=1}^{\infty} \mathbb{E}Y_n^2 < \infty$. More precisely the following estimate holds,*

$$\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 \le \frac{(\lambda + c)^2}{P(\sup_n |S_n| \le \lambda)} \text{ for all } \lambda > 0, \tag{12.20}$$

where as usual, $S_n := \sum_{j=1}^{n} Y_j$.

Remark 12.47. It follows from Eq. (12.20) that if $P(\sup_n |S_n| < \infty) > 0$, then $\sum_{j=1}^{\infty} \mathbb{E}Y_j^2 < \infty$ and hence by Kolmogorov's Theorem, $\sum_{j=1}^{\infty} Y_j = \lim_{n\to\infty} S_n$ exists a.s. and in particular, $P(\sup_n |S_n| < \infty)$.

Proof. Let $\lambda > 0$ and τ be the first time $|S_n| > \lambda$, i.e. let τ be the "stopping time" defined by,

$$\tau = \tau_\lambda := \inf \{n \ge 1 : |S_n| > \lambda\}.$$

As usual, $\tau = \infty$ if $\{n \ge 1 : |S_n| > \lambda\} = \emptyset$. Then for $N \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\left[S_N^2\right] &= \mathbb{E}\left[S_N^2 : \tau \le N\right] + \mathbb{E}\left[S_N^2 : \tau > N\right] \\ &\le \mathbb{E}\left[S_N^2 : \tau \le N\right] + \lambda^2 P[\tau > N]. \end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}\left[S_N^2 : \tau \le N\right] &= \sum_{j=1}^{N} \mathbb{E}\left[S_N^2 : \tau = j\right] = \sum_{j=1}^{N} \mathbb{E}\left[\left|S_j + S_N - S_j\right|^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 + 2S_j\left(S_N - S_j\right) + \left(S_N - S_j\right)^2 : \tau = j\right] \\
&= \sum_{j=1}^{N} \mathbb{E}\left[S_j^2 : \tau = j\right] + \sum_{j=1}^{N} \mathbb{E}\left[\left(S_N - S_j\right)^2\right] P\left[\tau = j\right] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[\left(S_{j-1} + Y_j\right)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] \sum_{j=1}^{N} P\left[\tau = j\right] \\
&\le \sum_{j=1}^{N} \mathbb{E}\left[\left(\lambda + c\right)^2 : \tau = j\right] + \mathbb{E}\left[S_N^2\right] P\left[\tau \le N\right] \\
&= \left[\left(\lambda + c\right)^2 + \mathbb{E}\left[S_N^2\right]\right] P\left[\tau \le N\right].
\end{aligned}$$

Putting this all together then gives,

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &\le \left[\left(\lambda + c\right)^2 + \mathbb{E}\left[S_N^2\right]\right] P\left[\tau \le N\right] + \lambda^2 P\left[\tau > N\right] \\
&\le \left[\left(\lambda + c\right)^2 + \mathbb{E}\left[S_N^2\right]\right] P\left[\tau \le N\right] + \left(\lambda + c\right)^2 P\left[\tau > N\right] \\
&= \left(\lambda + c\right)^2 + P\left[\tau \le N\right] \cdot \mathbb{E}\left[S_N^2\right]
\end{aligned}$$

form which it follows that

$$\begin{aligned}
\mathbb{E}\left[S_N^2\right] &\le \frac{\left(\lambda + c\right)^2}{1 - P\left[\tau \le N\right]} \le \frac{\left(\lambda + c\right)^2}{1 - P\left[\tau < \infty\right]} = \frac{\left(\lambda + c\right)^2}{P\left[\tau = \infty\right]} \\
&= \frac{\left(\lambda + c\right)^2}{P\left(\sup_n \left|S_n\right| \le \lambda\right)}.
\end{aligned}$$

Since S_n is convergent a.s., it follows that $P\left(\sup_n \left|S_n\right| < \infty\right) = 1$ and therefore,

$$\lim_{\lambda \uparrow \infty} P\left(\sup_n \left|S_n\right| < \lambda\right) = 1.$$

Hence for λ sufficiently large, $P\left(\sup_n \left|S_n\right| < \lambda\right) > 0$ ad we learn that

$$\sum_{j=1}^{\infty} \mathbb{E} Y_j^2 = \lim_{N \to \infty} \mathbb{E}\left[S_N^2\right] \le \frac{\left(\lambda + c\right)^2}{P\left(\sup_n \left|S_n\right| \le \lambda\right)} < \infty.$$

■

Lemma 12.48. *Suppose that* $\{Y_n\}_{n=1}^{\infty}$ *are independent random variables such that there exists* $c < \infty$ *such that* $|Y_n| \le c$ *a.s. for all* n*. If* $\sum_{n=1}^{\infty} Y_n$ *converges in* $\mathbb{R}$ *a.s. then* $\sum_{n=1}^{\infty} \mathbb{E} Y_n$ *converges as well.*

Proof. Let $(\Omega_0, \mathcal{B}_0, P_0)$ be the probability space that $\{Y_n\}_{n=1}^{\infty}$ is defined on and let

$$\Omega := \Omega_0 \times \Omega_0,\ \mathcal{B} := \mathcal{B}_0 \otimes \mathcal{B}_0, \text{ and } P := P_0 \otimes P_0.$$

Further let $Y_n'(\omega_1, \omega_2) := Y_n(\omega_1)$ and $Y_n''(\omega_1, \omega_2) := Y_n(\omega_2)$ and

$$Z_n(\omega_1, \omega_2) := Y_n'(\omega_1, \omega_2) - Y_n''(\omega_1, \omega_2) = Y_n(\omega_1) - Y_n(\omega_2).$$

Then $|Z_n| \le 2c$ a.s., $\mathbb{E}Z_n = 0$, and

$$\sum_{n=1}^{\infty} Z_n(\omega_1, \omega_2) = \sum_{n=1}^{\infty} Y_n(\omega_1) - \sum_{n=1}^{\infty} Y_n(\omega_2) \text{ exists}$$

for P a.e. (ω_1, ω_2). Hence it follows from Lemma 12.46 that

$$\begin{aligned} \infty > \sum_{n=1}^{\infty} \mathbb{E}Z_n^2 &= \sum_{n=1}^{\infty} \operatorname{Var}(Z_n) = \sum_{n=1}^{\infty} \operatorname{Var}(Y_n' - Y_n'') \\ &= \sum_{n=1}^{\infty} [\operatorname{Var}(Y_n') + \operatorname{Var}(Y_n'')] = 2\sum_{n=1}^{\infty} \operatorname{Var}(Y_n). \end{aligned}$$

Thus by Kolmogorov's convergence theorem, it follows that $\sum_{n=1}^{\infty}(Y_n - \mathbb{E}Y_n)$ is convergent. Since $\sum_{n=1}^{\infty} Y_n$ is a.s. convergent, we may conclude that $\sum_{n=1}^{\infty} \mathbb{E}Y_n$ is also convergent. ■

We are now ready to complete the proof of Theorem 12.12.

Proof. Our goal is to show if $\{X_n\}_{n=1}^{\infty}$ are independent random variables, then the random series, $\sum_{n=1}^{\infty} X_n$, is almost surely convergent iff for all $c > 0$ the following three series converge;

1. $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$, and
3. $\sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ converges.

Since $\sum_{n=1}^{\infty} X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and hence for every $c > 0$, $P(\{|X_n| \ge c \text{ i.o.}\}) = 0$. According the Borel zero one law this implies for every $c > 0$ that $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$. Given this, we now know that $\{X_n\}$ and $\left\{X_n^c := X_n 1_{|X_n| \le c}\right\}$ are tail equivalent for all $c > 0$ and in particular $\sum_{n=1}^{\infty} X_n^c$ is almost surely convergent for all $c > 0$. So according to Lemma 12.48 (with $Y_n = X_n^c$),

$$\sum_{n=1}^{\infty} \mathbb{E}X_n^c = \sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right) \text{ converges.}$$

Letting $Y_n := X_n^c - \mathbb{E}X_n^c$, we may now conclude that $\sum_{n=1}^{\infty} Y_n$ is almost surely convergent. Since $\{Y_n\}$ is uniformly bounded and $\mathbb{E}Y_n = 0$ for all n, an application of Lemma 12.46 allows us to conclude

$$\sum_{n=1}^{\infty} \mathrm{Var}\left(X_n 1_{|X_n| \leq c}\right) = \sum_{n=1}^{\infty} \mathbb{E} Y_n^2 < \infty.$$

■

13

Weak Convergence Results

Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables and X is another random variable (possibly defined on a different probability space). We would like to understand when, for large n, X_n and X have nearly the "same" distribution. Alternatively put, if we let $\mu_n(A) := P(X_n \in A)$ and $\mu(A) := P(X \in A)$, when is μ_n close to μ for large n. This is the question we will address in this chapter.

13.1 Total Variation Distance

Definition 13.1. *Let μ and ν be two probability measure on a measurable space, $(\Omega, \mathcal{B})$. The total variation distance, $d_{TV}(\mu, \nu)$, is defined as*

$$d_{TV}(\mu, \nu) := \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|.$$

Remark 13.2. The function, $\lambda : \mathcal{B} \to \mathbb{R}$ defined by, $\lambda(A) := \mu(A) - \nu(A)$ for all $A \in \mathcal{B}$, is an example of a "signed measure." For signed measures, one usually defines

$$\|\lambda\|_{TV} := \sup \left\{ \sum_{i=1}^{n} |\lambda(A_i)| : n \in \mathbb{N} \text{ and partitions, } \{A_i\}_{i=1}^{n} \subset \mathcal{B} \text{ of } \Omega \right\}.$$

You are asked to show in Exercise 13.1 below, that when $\lambda = \mu - \nu$, $d_{TV}(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_{TV}$.

Lemma 13.3 (Scheffé's Lemma). *Suppose that m is another positive measure on $(\Omega, \mathcal{B})$ such that there exists measurable functions, $f, g : \Omega \to [0, \infty)$, such that $d\mu = f dm$ and $d\nu = g dm$.*[1] *Then*

$$d_{TV}(\mu, \nu) = \frac{1}{2} \int_{\Omega} |f - g| \, dm.$$

Moreover, if $\{\mu_n\}_{n=1}^{\infty}$ is a sequence of probability measure of the form, $d\mu_n = f_n dm$ with $f_n : \Omega \to [0, \infty)$, and $f_n \to g$, m - a.e., then $d_{TV}(\mu_n, \nu) \to 0$ as $n \to \infty$.

[1] Fact: it is always possible to do this by taking $m = \mu + \nu$ for example.

Proof. Let $\lambda = \mu - \nu$ and $h := f - g : \Omega \to \mathbb{R}$ so that $d\lambda = hdm$. Since

$$\lambda(\Omega) = \mu(\Omega) - \nu(\Omega) = 1 - 1 = 0,$$

if $A \in \mathcal{B}$ we have

$$\lambda(A) + \lambda(A^c) = \lambda(\Omega) = 0.$$

In particular this shows $|\lambda(A)| = |\lambda(A^c)|$ and therefore,

$$\begin{aligned} |\lambda(A)| &= \frac{1}{2}\left[|\lambda(A)| + |\lambda(A^c)|\right] = \frac{1}{2}\left[\left|\int_A hdm\right| + \left|\int_{A^c} hdm\right|\right] \\ &\le \frac{1}{2}\left[\int_A |h|\, dm + \int_{A^c} |h|\, dm\right] = \frac{1}{2}\int_\Omega |h|\, dm. \end{aligned} \tag{13.1}$$

This shows

$$d_{TV}(\mu,\nu) = \sup_{A\in\mathcal{B}} |\lambda(A)| \le \frac{1}{2}\int_\Omega |h|\, dm.$$

To prove the converse inequality, simply take $A = \{h > 0\}$ (note $A^c = \{h \le 0\}$) in Eq. (13.1) to find

$$\begin{aligned} |\lambda(A)| &= \frac{1}{2}\left[\int_A hdm - \int_{A^c} hdm\right] \\ &= \frac{1}{2}\left[\int_A |h|\, dm + \int_{A^c} |h|\, dm\right] = \frac{1}{2}\int_\Omega |h|\, dm. \end{aligned}$$

For the second assertion, let $G_n := f_n + g$ and observe that $|f_n - g| \to 0$ $m-$a.e., $|f_n - g| \le G_n \in L^1(m)$, $G_n \to G := 2g$ a.e. and $\int_\Omega G_n dm = 2 \to 2 = \int_\Omega G dm$ and $n \to \infty$. Therefore, by the dominated convergence theorem 8.34,

$$\lim_{n\to\infty} d_{TV}(\mu_n, \nu) = \frac{1}{2}\lim_{n\to\infty}\int_\Omega |f_n - g|\, dm = 0.$$

■

For a concrete application of Scheffé's Lemma, see Proposition 13.35 below.

Corollary 13.4. *Let $\|h\|_\infty := \sup_{\omega\in\Omega} |h(\omega)|$ when $h : \Omega \to \mathbb{R}$ is a bounded random variable. Continuing the notation in Scheffé's lemma above, we have*

$$d_{TV}(\mu,\nu) = \frac{1}{2}\sup\left\{\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right| : \|h\|_\infty \le 1\right\}. \tag{13.2}$$

Consequently,

$$\left|\int_\Omega hd\mu - \int_\Omega hd\nu\right| \le 2d_{TV}(\mu,\nu)\cdot\|h\|_\infty \tag{13.3}$$

and in particular, for all bounded and measurable functions, $h : \Omega \to \mathbb{R}$,

$$\int_{\Omega} h d\mu_n \to \int_{\Omega} h d\nu \text{ if } d_{TV}(\mu_n, \nu) \to 0. \tag{13.4}$$

Proof. *We begin by observing that*

$$\left|\int_{\Omega} h d\mu - \int_{\Omega} h d\nu\right| = \left|\int_{\Omega} h(f-g)\, dm\right| \le \int_{\Omega} |h| \, |f-g| \, dm$$
$$\le \|h\|_{\infty} \int_{\Omega} |f-g| \, dm = 2 d_{TV}(\mu,\nu) \|h\|_{\infty}.$$

Moreover, from the proof of Scheffé's Lemma 13.3, we have

$$d_{TV}(\mu,\nu) = \frac{1}{2}\left|\int_{\Omega} h d\mu - \int_{\Omega} h d\nu\right|$$

when $h := 1_{f>g} - 1_{f\le g}$. *These two equations prove Eqs. (13.2) and (13.3) and the latter implies Eq. (13.4).* ■

Exercise 13.1. Under the hypothesis of Scheffé's Lemma 13.3, show

$$\|\mu - \nu\|_{TV} = \int_{\Omega} |f-g| \, dm = 2 d_{TV}(\mu,\nu).$$

Exercise 13.2. Suppose that Ω is a (at most) countable set, $\mathcal{B} := 2^{\Omega}$, and $\{\mu_n\}_{n=0}^{\infty}$ are probability measures on $(\Omega, \mathcal{B})$. Let $f_n(\omega) := \mu_n(\{\omega\})$ for $\omega \in \Omega$. Show

$$d_{TV}(\mu_n, \mu_0) = \frac{1}{2}\sum_{\omega\in\Omega} |f_n(\omega) - f_0(\omega)|$$

and $\lim_{n\to\infty} d_{TV}(\mu_n, \mu_0) = 0$ iff $\lim_{n\to\infty} \mu_n(\{\omega\}) = \mu_0(\{\omega\})$ for all $\omega \in \Omega$.

Notation 13.5 *Suppose that X and Y are random variables, let*

$$d_{TV}(X,Y) := d_{TV}(\mu_X, \mu_Y) = \sup_{A\in\mathcal{B}_{\mathbb{R}}} |P(X\in A) - P(Y\in A)|,$$

where $\mu_X = P\circ X^{-1}$ and $\mu_Y = P \circ Y^{-1}$.

13.2 Weak Convergence

Example 13.6. Suppose that $P\left(X_n = \frac{i}{n}\right) = \frac{1}{n}$ for $i \in \{1,2,\dots,n\}$ so that X_n is a discrete "approximation" to the uniform distribution, i.e. to U where $P(U\in A) = m(A\cap[0,1])$ for all $A \in \mathcal{B}_{\mathbb{R}}$. If we let $A_n = \left\{\frac{i}{n} : i = 1,2,\dots,n\right\}$, then $P(X_n \in A_n) = 1$ while $P(U\in A_n) = 0$. Therefore, it follows that $d_{TV}(X_n, U) = 1$ for all n.[2]

[2] More generally, if μ and ν are two probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$ while ν concentrates on a countable set, then $d_{TF}(\mu,\nu) = 1$.

Nevertheless we would like X_n to be close to U in distribution. Let us observe that if we let $F_n(y) := P(X_n \le y)$ and $F(y) := P(U \le y)$, then

$$F_n(y) = P(X_n \le y) = \frac{1}{n}\#\left\{i \in \{1,2,\dots,n\} : \frac{i}{n} \le y\right\}$$

and

$$F(y) := P(U \le y) = (y \wedge 1) \vee 0.$$

From these formula, it easily follows that $F(y) = \lim_{n\to\infty} F_n(y)$ for all $y \in \mathbb{R}$. This suggest that we should say that X_n converges in distribution to X iff $P(X_n \le y) \to P(X \le y)$ for all $y \in \mathbb{R}$. However, the next simple example shows this definition is also too restrictive.

Example 13.7. Suppose that $P(X_n = 1/n) = 1$ for all n and $P(X_0 = 0) = 1$. Then it is reasonable to insist that X_n converges of X_0 in distribution. However, $F_n(y) = 1_{y\ge 1/n} \to 1_{y\ge 0} = F_0(y)$ for all $y \in \mathbb{R}$ **except** for $y = 0$. Observe that y is the only point of discontinuity of F_0.

Notation 13.8 *Let (X,d) be a metric space, $f : X \to \mathbb{R}$ be a function. The set of $x \in X$ where f is continuous (discontinuous) at x will be denoted by $\mathcal{C}(f)$ ($\mathcal{D}(f)$).*

Observe that if $F : \mathbb{R} \to [0,1]$ is a non-decreasing function, then $\mathcal{C}(F)$ is at most countable. To see this, suppose that $\varepsilon > 0$ is given and let $\mathcal{C}_\varepsilon := \{y \in \mathbb{R} : F(y+) - F(y-) \ge \varepsilon\}$. If $y < y'$ with $y, y' \in \mathcal{C}_\varepsilon$, then $F(y+) < F(y'-)$ and $(F(y-), F(y+))$ and $(F(y'-), F(y'+))$ are disjoint intervals of length greater that ε. Hence it follows that

$$1 = m([0,1]) \ge \sum_{y\in\mathcal{C}_\varepsilon} m((F(y-), F(y+))) \ge \varepsilon \cdot \#(\mathcal{C}_\varepsilon)$$

and hence that $\#(\mathcal{C}_\varepsilon) \le \varepsilon^{-1} < \infty$. Therefore $\mathcal{C} := \cup_{k=1}^\infty \mathcal{C}_{1/k}$ is at most countable.

Definition 13.9. *Let $\{F, F_n : n = 1,2,\dots\}$ be a collection of right continuous non-increasing functions from $\mathbb{R}$ to $[0,1]$ and by abuse of notation let us also denote the associated measures, μ_F and μ_{F_n} by F and F_n respectively. Then*

1. *F_n converges to F **vaguely** and write, $F_n \xrightarrow{v} F$, iff $F_n((a,b]) \to F((a,b])$ for all $a, b \in \mathcal{C}(F)$.*
2. *F_n converges to F **weakly** and write, $F_n \xrightarrow{w} F$, iff $F_n(x) \to F(x)$ for all $x \in \mathcal{C}(F)$.*
3. *We say F is **proper**, if F is a distribution function of a probability measure, i.e. if $F(\infty) = 1$ and $F(-\infty) = 0$.*

Example 13.10. If X_n and U are as in Example 13.6 and $F_n(y) := P(X_n \le y)$ and $F(y) := P(Y \le y)$, then $F_n \xrightarrow{v} F$ and $F_n \xrightarrow{w} F$.

Lemma 13.11. *Let $\{F, F_n : n = 1, 2, \dots\}$ be a collection of proper distribution functions. Then $F_n \overset{v}{\to} F$ iff $F_n \overset{w}{\to} F$. In the case where F_n and F are proper and $F_n \overset{w}{\to} F$, we will write $F_n \implies F$.*

Proof. If $F_n \overset{w}{\to} F$, then $F_n((a,b]) = F_n(b) - F_n(a) \to F(b) - F(a) = F((a,b])$ for all $a, b \in \mathcal{C}(F)$ and therefore $F_n \overset{v}{\to} F$. So now suppose $F_n \overset{v}{\to} F$ and let $a < x$ with $a, x \in \mathcal{C}(F)$. Then

$$F(x) = F(a) + \lim_{n\to\infty} [F_n(x) - F_n(a)] \le F(a) + \liminf_{n\to\infty} F_n(x).$$

Letting $a \downarrow -\infty$, using the fact that F is proper, implies

$$F(x) \le \liminf_{n\to\infty} F_n(x).$$

Likewise,

$$F(x) - F(a) = \lim_{n\to\infty} [F_n(x) - F_n(a)] \ge \limsup_{n\to\infty} [F_n(x) - 1] = \limsup_{n\to\infty} F_n(x) - 1$$

which upon letting $a \uparrow \infty$, (so $F(a) \uparrow 1$) allows us to conclude,

$$F(x) \ge \limsup_{n\to\infty} F_n(x).$$

■

Definition 13.12. *A sequence of random variables, $\{X_n\}_{n=1}^\infty$ is said to* ***converge weakly*** *or to* ***converge in distribution*** *to a random variable X (written $X_n \implies X$) iff $F_n(y) := P(X_n \le y) \implies F(y) := P(X \le y)$.*

Example 13.13 (Central Limit Theorem). The central limit theorem (see the next chapter) states; if $\{X_n\}_{n=1}^\infty$ are i.i.d. $L^2(P)$ random variables with $\mu := \mathbb{E}X_1$ and $\sigma^2 = \mathrm{Var}(X_1)$, then

$$\frac{S_n - n\mu}{\sqrt{n}} \implies N(0,\sigma) \overset{d}{=} \sigma N(0,1).$$

Written out explicitly we find

$$\begin{aligned}\lim_{n\to\infty} P\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) &= P(a < N(0,1) \le b) \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx\end{aligned}$$

or equivalently put

$$\lim_{n\to\infty} P\left(n\mu + \sigma\sqrt{n}a < S_n \le n\mu + \sigma\sqrt{n}b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

More intuitively, we have

$$S_n \overset{d}{\cong} n\mu + \sqrt{n}\sigma N(0,1) \overset{d}{=} N\left(n\mu, n\sigma^2\right).$$

Lemma 13.14. *Suppose X is a random variable, $\{c_n\}_{n=1}^\infty \subset \mathbb{R}$, and $X_n = X + c_n$. If $c := \lim_{n\to\infty} c_n$ exists, then $X_n \implies X + c$.*

Proof. Let $F(x) := P(X \leq x)$ and

$$F_n(x) := P(X_n \leq x) = P(X + c_n \leq x) = F(x - c_n).$$

Clearly, if $c_n \to c$ as $n \to \infty$, then for all $x \in \mathcal{C}(F(\cdot - c))$ we have $F_n(x) \to F(x-c)$. Since $F(x-c) = P(X + c \leq x)$, we see that $X_n \implies X + c$. Observe that $F_n(x) \to F(x-c)$ only for $x \in \mathcal{C}(F(\cdot - c))$ but this is sufficient to assert $X_n \implies X + c$. ∎

Example 13.15. Suppose that $P(X_n = n) = 1$ for all n, then $F_n(y) = 1_{y \geq n} \to 0 = F(y)$ as $n \to \infty$. Notice that F is not a distribution function because all of the mass went off to $+\infty$. Similarly, if we suppose, $P(X_n = \pm n) = \frac{1}{2}$ for all n, then $F_n = \frac{1}{2} 1_{[-n,n)} + 1_{[n,\infty)} \to \frac{1}{2} = F(y)$ as $n \to \infty$. Again, F is not a distribution function on $\mathbb{R}$ since half the mass went to $-\infty$ while the other half went to $+\infty$.

Example 13.16. Suppose X is a non-zero random variables such that $X \overset{d}{=} -X$, then $X_n := (-1)^n X \overset{d}{=} X$ for all n and therefore, $X_n \implies X$ as $n \to \infty$. On the other hand, X_n does not converge to X almost surely or in probability.

The next theorem summarizes a number of useful equivalent characterizations of weak convergence. (The reader should compare Theorem 13.17 with Corollary 13.4.) In this theorem we will write $BC(\mathbb{R})$ for the bounded continuous functions, $f : \mathbb{R} \to \mathbb{R}$ (or $f : \mathbb{R} \to \mathbb{C}$) and $C_c(\mathbb{R})$ for those $f \in C(\mathbb{R})$ which have compact support, i.e. $f(x) \equiv 0$ if $|x|$ is sufficiently large.

Theorem 13.17. *Suppose that $\{\mu_n\}_{n=0}^\infty$ is a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ and for each n, let $F_n(y) := \mu_n((-\infty, y])$ be the (proper) distribution function associated to μ_n. Then the following are equivalent.*

1. *For all $f \in BC(\mathbb{R})$,*
$$\int_\mathbb{R} f d\mu_n \to \int_\mathbb{R} f d\mu_0 \text{ as } n \to \infty. \tag{13.5}$$
2. *Eq. (13.5) holds for all $f \in BC(\mathbb{R})$ which are uniformly continuous.*
3. *Eq. (13.5) holds for all $f \in C_c(\mathbb{R})$.*
4. *$F_n \implies F$.*
5. *There exists a probability space $(\Omega, \mathcal{B}, P)$ and random variables, Y_n, on this space such that $P \circ Y_n^{-1} = \mu_n$ for all n and $Y_n \to Y_0$ a.s.*

Proof. Clearly 1. $\implies$ 2. $\implies$ 3. and 5. $\implies$ 1. by the dominated convergence theorem. Indeed, we have

$$\int_\mathbb{R} f d\mu_n = \mathbb{E}[f(Y_n)] \overset{\text{D.C.T.}}{\to} \mathbb{E}[f(Y)] = \int_\mathbb{R} f d\mu_0$$

for all $f \in BC(\mathbb{R})$. Therefore it suffices to prove 3. $\Longrightarrow$ 4. and 4. $\Longrightarrow$ 5. The proof of 4. $\Longrightarrow$ 5. will be the content of Skorohod's Theorem 13.28 below. Given Skorohod's Theorem, we will now complete the proof.

(3. $\Longrightarrow$ 4.) Let $-\infty < a < b < \infty$ with $a, b \in \mathcal{C}(F_0)$ and for $\varepsilon > 0$, let $f_\varepsilon(x) \geq 1_{(a,b]}$ and $g_\varepsilon(x) \leq 1_{(a,b]}$ be the functions in $C_c(\mathbb{R})$ pictured in Figure 13.1. Then

$$\limsup_{n\to\infty} \mu_n((a,b]) \leq \limsup_{n\to\infty} \int_{\mathbb{R}} f_\varepsilon d\mu_n = \int_{\mathbb{R}} f_\varepsilon d\mu_0 \tag{13.6}$$

and

$$\liminf_{n\to\infty} \mu_n((a,b]) \geq \liminf_{n\to\infty} \int_{\mathbb{R}} g_\varepsilon d\mu_n = \int_{\mathbb{R}} g_\varepsilon d\mu_0. \tag{13.7}$$

Since $f_\varepsilon \to 1_{[a,b]}$ and $g_\varepsilon \to 1_{(a,b)}$ as $\varepsilon \downarrow 0$, we may use the dominated convergence theorem to pass to the limit as $\varepsilon \downarrow 0$ in Eqs. (13.6) and (13.7) to conclude,

$$\limsup_{n\to\infty} \mu_n((a,b]) \leq \mu_0([a,b]) = \mu_0((a,b])$$

and

$$\liminf_{n\to\infty} \mu_n((a,b]) \geq \mu_0((a,b)) = \mu_0((a,b]),$$

where the second equality in each of the equations holds because a and b are points of continuity of F_0. Hence we have shown that $\lim_{n\to\infty} \mu_n((a,b])$ exists and is equal to $\mu_0((a,b])$.

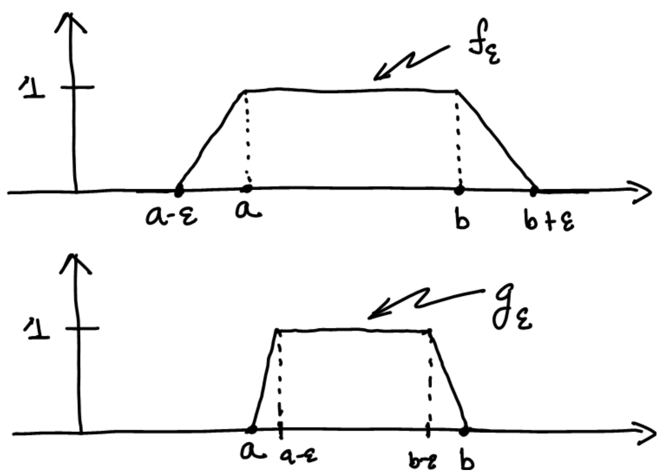


Fig. 13.1. The picture definition of the trapezoidal functions, f_ε and g_ε.

■

Corollary 13.18. *Suppose that* $\{X_n\}_{n=0}^{\infty}$ *is a sequence of random variables, such that* $X_n \xrightarrow{P} X_0$*, then* $X_n \implies X_0$. *(Recall that example 13.16 shows the converse is in general false.)*

Proof. Let $g \in BC(\mathbb{R})$, then by Corollary 11.9, $g(X_n) \xrightarrow{P} g(X_0)$ and since g is bounded, we may apply the dominated convergence theorem (see Corollary 11.8) to conclude that $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X_0)]$. ∎

Lemma 13.19. *Suppose* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of random variables on a common probability space and* $c \in \mathbb{R}$. *Then* $X_n \implies c$ *iff* $X_n \xrightarrow{P} c$.

Proof. Recall that $X_n \xrightarrow{P} c$ iff for all $\varepsilon > 0$, $P(|X_n - c| > \varepsilon) \to 0$. Since

$$\{|X_n - c| > \varepsilon\} = \{X_n > c + \varepsilon\} \cup \{X_n < c - \varepsilon\}$$

it follows $X_n \xrightarrow{P} c$ iff $P(X_n > x) \to 0$ for all $x > c$ and $P(X_n < x) \to 0$ for all $x < c$. These conditions are also equivalent to $P(X_n \le x) \to 1$ for all $x > c$ and $P(X_n \le x) \le P(X_n \le x') \to 0$ for all $x < c$ (where $x < x' < c$). So $X_n \xrightarrow{P} c$ iff

$$\lim_{n\to\infty} P(X_n \le x) = \begin{cases} 0 \text{ if } x < c \\ 1 \text{ if } x > c \end{cases} = F(x)$$

where $F(x) = P(c \le x) = 1_{x \ge c}$. Since $\mathcal{C}(F) = \mathbb{R} \setminus \{c\}$, we have shown $X_n \xrightarrow{P} c$ iff $X_n \implies c$. ∎

We end this section with a few more equivalent characterizations of weak convergence. The combination of Theorem 13.17 and 13.20 is often called the Portmanteau Theorem.

Theorem 13.20 (The Portmanteau Theorem). *Suppose* $\{F_n\}_{n=0}^{\infty}$ *are proper distribution functions. By abuse of notation, we will denote* $\mu_{F_n}(A)$ *simply by* $F_n(A)$ *for all* $A \in \mathcal{B}_{\mathbb{R}}$. *Then the following are equivalent.*

1. $F_n \implies F_0$.
2. $\liminf_{n\to\infty} F_n(U) \ge F_0(U)$ *for open subsets,* $U \subset \mathbb{R}$.
3. $\limsup_{n\to\infty} F_n(C) \le F_0(C)$ *for all closed subsets,* $C \subset \mathbb{R}$.
4. $\lim_{n\to\infty} F_n(A) = F_0(A)$ *for all* $A \in \mathcal{B}_{\mathbb{R}}$ *such that* $F_0(\partial A) = 0$.

Proof. (1. $\implies$ 2.) By Theorem 13.28 we may choose random variables, Y_n, such that $P(Y_n \le y) = F_n(y)$ for all $y \in \mathbb{R}$ and $n \in \mathbb{N}$ and $Y_n \to Y_0$ a.s. as $n \to \infty$. Since U is open, it follows that

$$1_U(Y) \le \liminf_{n\to\infty} 1_U(Y_n) \text{ a.s.}$$

and so by Fatou's lemma,

$$\begin{aligned} F(U) &= P(Y \in U) = \mathbb{E}[1_U(Y)] \\ &\le \liminf_{n\to\infty} \mathbb{E}[1_U(Y_n)] = \liminf_{n\to\infty} P(Y_n \in U) = \liminf_{n\to\infty} F_n(U). \end{aligned}$$

(2. $\iff$ 3.) This follows from the observations: 1) $C \subset \mathbb{R}$ is closed iff $U := C^c$ is open, 2) $F(U) = 1 - F(C)$, and 3) $\liminf_{n\to\infty}(-F_n(C)) = -\limsup_{n\to\infty} F_n(C)$.

(2. and 3. $\iff$ 4.) If $F_0(\partial A) = 0$, then $A^o \subset A \subset \bar{A}$ with $F_0\left(\bar{A} \setminus A^o\right) = 0$. Therefore

$$F_0(A) = F_0(A^o) \leq \liminf_{n\to\infty} F_n(A^o) \leq \limsup_{n\to\infty} F_n\left(\bar{A}\right) \leq F_0\left(\bar{A}\right) = F_0(A).$$

(4. $\implies$ 1.) Let $a, b \in \mathcal{C}(F_0)$ and take $A := (a, b]$. Then $F_0(\partial A) = F_0(\{a, b\}) = 0$ and therefore, $\lim_{n\to\infty} F_n((a, b]) = F_0((a, b])$, i.e. $F_n \implies F_0$. ■

Exercise 13.3. Suppose that F is a continuous proper distribution function. Show,

1. $F : \mathbb{R} \to [0, 1]$ is uniformly continuous.
2. If $\{F_n\}_{n=1}^\infty$ is a sequence of distribution functions converging weakly to F, then F_n converges to F uniformly on $\mathbb{R}$, i.e.

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F(x) - F_n(x)| = 0.$$

In particular, it follows that

$$\begin{aligned}\sup_{a<b} |\mu_F((a, b]) - \mu_{F_n}((a, b])| &= \sup_{a<b} |F(b) - F(a) - (F_n(b) - F_n(a))| \\ &\leq \sup_b |F(b) - F_n(b)| + \sup_a |F_n(a) - F_n(a)| \\ &\to 0 \text{ as } n \to \infty.\end{aligned}$$

Hints for part 2. Given $\varepsilon > 0$, show that there exists, $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_n = \infty$, such that $|F(\alpha_{i+1}) - F(\alpha_i)| \leq \varepsilon$ for all i. Now show, for $x \in [\alpha_i, \alpha_{i+1})$, that

$$|F(x) - F_n(x)| \leq (F(\alpha_{i+1}) - F(\alpha_i)) + |F(\alpha_i) - F_n(\alpha_i)| + (F_n(\alpha_{i+1}) - F_n(\alpha_i)).$$

13.3 "Derived" Weak Convergence

Lemma 13.21. *Let (X, d) be a metric space, $f : X \to \mathbb{R}$ be a function, and $\mathcal{D}(f)$ be the set of $x \in X$ where f is discontinuous at x. Then $\mathcal{D}(f)$ is a Borel measurable subset of X.*

Proof. For $x \in X$ and $\delta > 0$, let $B_x(\delta) = \{y \in X : d(x, y) < \delta\}$. Given $\delta > 0$, let $f^\delta : X \to \mathbb{R} \cup \{\infty\}$ be defined by,

$$f^\delta(x) := \sup_{y\in B_x(\delta)} f(y).$$

We will begin by showing f^δ is **lower semi-continuous**, i.e. $\{f^\delta \le a\}$ is closed (or equivalently $\{f^\delta > a\}$ is open) for all $a \in \mathbb{R}$. Indeed, if $f^\delta(x) > a$, then there exists $y \in B_x(\delta)$ such that $f(y) > a$. Since this y is in $B_{x'}(\delta)$ whenever $d(x,x') < \delta - d(x,y)$ (because then, $d(x',y) \le d(x,y) + d(x,x') < \delta$) it follows that $f^\delta(x') > a$ for all $x' \in B_x(\delta - d(x,y))$. This shows $\{f^\delta > a\}$ is open in X.

We similarly define $f_\delta : X \to \mathbb{R} \cup \{-\infty\}$ by

$$f_\delta(x) := \inf_{y \in B_x(\delta)} f(y).$$

Since $f_\delta = -(-f)^\delta$, it follows that

$$\{f_\delta \ge a\} = \left\{(-f)^\delta \le -a\right\}$$

is closed for all $a \in \mathbb{R}$, i.e. f_δ is **upper semi-continuous**. Moreover, $f_\delta \le f \le f^\delta$ for all $\delta > 0$ and $f^\delta \downarrow f^0$ and $f_\delta \uparrow f_0$ as $\delta \downarrow 0$, where $f_0 \le f \le f^0$ and $f_0 : X \to \mathbb{R} \cup \{-\infty\}$ and $f^0 : X \to \mathbb{R} \cup \{\infty\}$ are measurable functions. The proof is now complete since it is easy to see that

$$\mathcal{D}(f) = \left\{f^0 > f_0\right\} = \left\{f^0 - f_0 \neq 0\right\} \in \mathcal{B}_X.$$

■

Remark 13.22. Suppose that $x_n \to x$ with $x \in \mathcal{C}(f) := \mathcal{D}(f)^c$. Then $f(x_n) \to f(x)$ as $n \to \infty$.

Theorem 13.23 (Continuous Mapping Theorem). *Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable functions. If $X_n \implies X_0$ and $P(X_0 \in \mathcal{D}(f)) = 0$, then $f(X_n) \implies f(X_0)$. If in addition, f is bounded, $\mathbb{E}f(X_n) \to \mathbb{E}f(X_0)$.*

Proof. Let $\{Y_n\}_{n=0}^\infty$ be random variables on some probability space as in Theorem 13.28. For $g \in BC(\mathbb{R})$ we observe that $\mathcal{D}(g \circ f) \subset \mathcal{D}(f)$ and therefore,

$$P(Y_0 \in \mathcal{D}(g \circ f)) \le P(Y_0 \in \mathcal{D}(f)) = P(X_0 \in \mathcal{D}(f)) = 0.$$

Hence it follows that $g \circ f \circ Y_n \to g \circ f \circ Y_0$ a.s. So an application of the dominated convergence theorem (see Corollary 11.8) implies

$$\mathbb{E}[g(f(X_n))] = \mathbb{E}[g(f(Y_n))] \to \mathbb{E}[g(f(Y_0))] = \mathbb{E}[g(f(X_0))]. \tag{13.8}$$

This proves the first assertion. For the second assertion we take $g(x) = (x \wedge M) \vee (-M)$ in Eq. (13.8) where M is a bound on $|f|$. ■

Theorem 13.24 (Slutzky's Theorem). *Suppose that $X_n \implies X$ and $Y_n \xrightarrow{P} c$ where c is a constant. Then $(X_n, Y_n) \implies (X, c)$ in the sense that $\mathbb{E}[f(X_n, Y_n)] \to \mathbb{E}[f(X, c)]$ for all $f \in BC(\mathbb{R}^2)$. In particular, by taking $f(x,y) = g(x+y)$ and $f(x,y) = g(x \cdot y)$ with $g \in BC(\mathbb{R})$, we learn $X_n + Y_n \implies X + c$ and $X_n \cdot Y_n \implies X \cdot c$ respectively.*

Proof. First suppose that $f \in C_c\left(\mathbb{R}^2\right)$, and for $\varepsilon > 0$, let $\delta := \delta(\varepsilon)$ be chosen so that

$$|f(x,y) - f(x',y')| \le \varepsilon \text{ if } \|(x,y) - (x',y')\| \le \delta.$$

Then

$$\begin{aligned}|\mathbb{E}[f(X_n,Y_n) - f(X_n,c)]| &\le \mathbb{E}[|f(X_n,Y_n) - f(X_n,c)| : |Y_n - c| \le \delta] \\ &\quad + \mathbb{E}[|f(X_n,Y_n) - f(X_n,c)| : |Y_n - c| > \delta] \\ &\le \varepsilon + 2MP(|Y_n - c| > \delta) \to \varepsilon \text{ as } n \to \infty,\end{aligned}$$

where $M = \sup|f|$. Since, $X_n \implies X$, we know $\mathbb{E}[f(X_n,c)] \to \mathbb{E}[f(X,c)]$ and hence we have shown,

$$\begin{aligned}&\limsup_{n\to\infty} |\mathbb{E}[f(X_n,Y_n) - f(X,c)]| \\ &\quad \le \limsup_{n\to\infty} |\mathbb{E}[f(X_n,Y_n) - f(X_n,c)]| + \limsup_{n\to\infty} |\mathbb{E}[f(X_n,c) - f(X,c)]| \le \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we learn that $\lim_{n\to\infty} \mathbb{E}f(X_n,Y_n) = \mathbb{E}f(X,c)$.

Now suppose $f \in BC\left(\mathbb{R}^2\right)$ with $f \ge 0$ and let $\varphi_k(x,y) \in [0,1]$ be continuous functions with compact support such that $\varphi_k(x,y) = 1$ if $|x| \vee |y| \le k$ and $\varphi_k(x,y) \uparrow 1$ as $k \to \infty$. Then applying what we have just proved to $f_k := \varphi_k f$, we find

$$\mathbb{E}[f_k(X,c)] = \lim_{n\to\infty} \mathbb{E}[f_k(X_n,Y_n)] \le \liminf_{n\to\infty} \mathbb{E}[f(X_n,Y_n)].$$

Letting $k \to \infty$ in this inequality then implies that

$$\mathbb{E}[f(X,c)] \le \liminf_{n\to\infty} \mathbb{E}[f(X_n,Y_n)].$$

This inequality with f replaced by $M - f \ge 0$ then shows,

$$M - \mathbb{E}[f(X,c)] \le \liminf_{n\to\infty} \mathbb{E}[M - f(X_n,Y_n)] = M - \limsup_{n\to\infty} \mathbb{E}[f(X_n,Y_n)].$$

Hence we have shown,

$$\limsup_{n\to\infty} \mathbb{E}[f(X_n,Y_n)] \le \mathbb{E}[f(X,c)] \le \liminf_{n\to\infty} \mathbb{E}[f(X_n,Y_n)]$$

and therefore $\lim_{n\to\infty} \mathbb{E}[f(X_n,Y_n)] = \mathbb{E}[f(X,c)]$ for all $f \in BC\left(\mathbb{R}^2\right)$ with $f \ge 0$. This completes the proof since any $f \in BC\left(\mathbb{R}^2\right)$ may be written as a difference of its positive and negative parts. ■

Theorem 13.25 (δ − method). *Suppose that $\{X_n\}_{n=1}^\infty$ are random variables, $b \in \mathbb{R}$, $a_n \in \mathbb{R}\setminus\{0\}$ with $\lim_{n\to\infty} a_n = 0$, and*

$$\frac{X_n - b}{a_n} \implies Z.$$

If $g : \mathbb{R} \to \mathbb{R}$ be a measurable function which is differentiable at b, then

$$\frac{g(X_n) - g(b)}{a_n} \implies g'(b)\, Z.$$

Proof. Observe that

$$X_n - b = a_n \frac{X_n - b}{a_n} \implies 0 \cdot Z = 0$$

so that $X_n \implies b$ and hence $X_n \xrightarrow{P} b$. By definition of the derivative of g at b, we have

$$g(x + \Delta) = g(b) + g'(b)\Delta + \varepsilon(\Delta)\Delta$$

where $\varepsilon(\Delta) \to 0$ as $\Delta \to 0$. Let Y_n and Y be random variables on a fixed probability space such that $Y_n \stackrel{d}{=} \frac{X_n - b}{a_n}$ and $Y \stackrel{d}{=} Z$ with $Y_n \to Y$ a.s. Then $X_n \stackrel{d}{=} a_n Y_n + b$, so that

$$\begin{aligned}\frac{g(X_n) - g(b)}{a_n} \stackrel{d}{=} \frac{g(a_n Y_n + b) - g(b)}{a_n} &= g'(b) Y_n + \frac{a_n Y_n \varepsilon(a_n Y_n)}{a_n} \\ &= g'(b) Y_n + Y_n \varepsilon(a_n Y_n) \to g'(b) Y \text{ a.s.}\end{aligned}$$

This completes the proof since $g'(b) Y \stackrel{d}{=} g'(b) Z$. ■

Example 13.26. Suppose that $\{U_n\}_{n=1}^\infty$ are i.i.d. random variables which are uniformly distributed on $[0,1]$ and let $Y_n := \prod_{j=1}^n U_j^{\frac{1}{n}}$. Our goal is to find a_n and b_n such that $\frac{Y_n - b_n}{a_n}$ is weakly convergent to a non-constant random variable. To this end, let

$$X_n := \ln Y_n = \frac{1}{n} \sum_{j=1}^n \ln U_j.$$

By the strong law of large numbers,

$$\lim_{n\to\infty} X_n \stackrel{a.s.}{=} \mathbb{E}[\ln U_1] = \int_0^1 \ln x dx = [x \ln x - x]_0^1 = -1$$

and therefore, $\lim_{n\to\infty} Y_n \stackrel{a.s.}{=} e^{-1}$.

Let us further observe that

$$\mathbb{E}\left[\ln^2 U_1\right] = \int_0^1 \ln^2 x dx = 2$$

so that $\operatorname{Var}(\ln U_1) = 2 - (-1)^2 = 1$. Hence by the central limit theorem,

$$\frac{X_n - (-1)}{\frac{1}{\sqrt{n}}} = \sqrt{n}(X_n + 1) \implies N(0,1).$$

Therefore the δ – method implies,

$$\frac{g(X_n) - g(-1)}{\frac{1}{\sqrt{n}}} \implies g'(-1) N(0,1).$$

Taking $g(x) := e^x$ using $g(X_n) = e^{X_n} = Y_n$, then implies

$$\frac{Y_n - e^{-1}}{\frac{1}{\sqrt{n}}} \implies e^{-1} N(0,1) \stackrel{d}{=} N\left(0, e^{-2}\right).$$

Hence we have shown,

$$\sqrt{n}\left[\prod_{j=1}^{n} U_j^{\frac{1}{n}} - e^{-1}\right] \implies N\left(0, e^{-2}\right).$$

Exercise 13.4. Given a function, $f : X \to \mathbb{R}$ and a point $x \in X$, let

$$\liminf_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \inf_{y \in B'_x(\delta)} f(y) \text{ and} \tag{13.9}$$

$$\limsup_{y \to x} f(y) := \lim_{\varepsilon \downarrow 0} \sup_{y \in B'_x(\delta)} f(y), \tag{13.10}$$

where

$$B'_x(\delta) := \{y \in X : 0 < d(x,y) < \delta\}.$$

Show f is lower (upper) semi-continuous iff $\liminf_{y \to x} f(y) \geq f(x)$ $\left(\limsup_{y \to x} f(y) \leq f(x)\right)$ for all $x \in X$.

Solution to Exercise (13.4). Suppose Eq. (13.9) holds, $a \in \mathbb{R}$, and $x \in X$ such that $f(x) > a$. Since,

$$\lim_{\varepsilon \downarrow 0} \inf_{y \in B'_x(\delta)} f(y) = \liminf_{y \to x} f(y) \geq f(x) > a,$$

it follows that $\inf_{y \in B'_x(\delta)} f(y) > a$ for some $\delta > 0$. Hence we may conclude that $B_x(\delta) \subset \{f > a\}$ which shows $\{f > a\}$ is open.

Conversely, suppose now that $\{f > a\}$ is open for all $a \in \mathbb{R}$. Given $x \in X$ and $a < f(x)$, there exists $\delta > 0$ such that $B_x(\delta) \subset \{f > a\}$. Hence it follows that $\liminf_{y \to x} f(y) \geq a$ and then letting $a \uparrow f(x)$ then implies $\liminf_{y \to x} f(y) \geq f(x)$.

13.4 Skorohod and the Convergence of Types Theorems

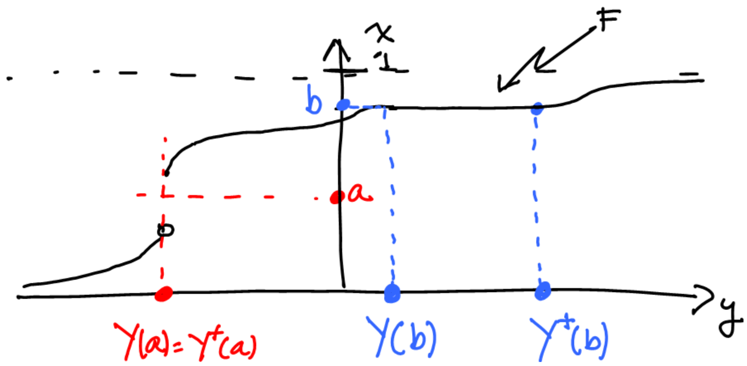
Notation 13.27 *Given a proper distribution function, $F : \mathbb{R} \to [0,1]$, let $Y = F^{\leftarrow} : (0,1) \to \mathbb{R}$ be the function defined by*

$$Y(x) = F^{\leftarrow}(x) = \sup\{y \in \mathbb{R} : F(y) < x\}.$$

Similarly, let

$$Y^+(x) := \inf\{y \in \mathbb{R} : F(y) > x\}.$$

We will need the following simple observations about Y and Y^+ which are easily understood from Figure 13.4.



1. $Y(x) \le Y^+(x)$ and $Y(x) < Y^+(x)$ iff x is the height of a "flat spot" of F.
2. The set, $E := \{x \in (0,1) : Y(x) < Y^+(x)\}$, of flat spot heights is at most countable. This is because, $\{(Y(x), Y^+(x))\}_{x\in E}$ is a collection of pairwise disjoint intervals which is necessarily countable. (Each such interval contains a rational number.)
3. The following inequality holds,
$$F(Y(x)-) \le x \le F(Y(x)) \text{ for all } x \in (0,1). \tag{13.11}$$
Indeed, if $y > Y(x)$, then $F(y) \ge x$ and by right continuity of F it follows that $F(Y(x)) \ge x$. Similarly, if $y < Y(x)$, then $F(y) < x$ and hence $F(Y(x)-) \le x$.
4. $\{x \in (0,1) : Y(x) \le y_0\} = (0, F(y_0)] \cap (0,1)$. To prove this assertion first suppose that $Y(x) \le y_0$, then according to Eq. (13.11) we have $x \le F(Y(x)) \le F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely, if $x \in (0,1)$ and $x \le F(y_0)$, then $Y(x) \le y_0$ by definition of Y.
5. As a consequence of item 4. we see that Y is $\mathcal{B}_{(0,1)}/\mathcal{B}_{\mathbb{R}}$ – measurable and $m \circ Y^{-1} = F$, where m is Lebesgue measure on $((0,1), \mathcal{B}_{(0,1)})$.

Theorem 13.28 (Baby Skorohod Theorem). *Suppose that $\{F_n\}_{n=0}^\infty$ is a collection of distribution functions such that $F_n \implies F_0$. Then there exists a probability space, $(\Omega, \mathcal{B}, P)$ and random variables, $\{Y_n\}_{n=1}^\infty$ such that $P(Y_n \le y) = F_n(y)$ for all $n \in \mathbb{N} \cup \{\infty\}$ and $\lim_{n\to\infty} F_n^{\leftarrow} = \lim_{n\to\infty} Y_n = Y = F^{\leftarrow}$ a.s.*

Proof. We will take $\Omega := (0,1)$, $\mathcal{B} = \mathcal{B}_{(0,1)}$, and $P = m$ – Lebesgue measure on Ω and let $Y_n := F_n^{\leftarrow}$ and $Y := F_0^{\leftarrow}$ as in Notation 13.27. Because of the above comments, $P(Y_n \le y) = F_n(y)$ and $P(Y \le y) = F_0(y)$ for all $y \in \mathbb{R}$. So in order to finish the proof it suffices to show, $Y_n(x) \to Y(x)$ for all $x \notin E$, where E is the countable null set defined as above, $E := \{x \in (0,1) : Y(x) < Y^+(x)\}$.

We now suppose $x \notin E$. If $y \in \mathcal{C}(F_0)$ with $y < Y(x)$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) < x$ and in particular, $F_n(y) < x$ for almost all n. This implies that $Y_n(x) \geq y$ for a.a. n and hence that $\liminf_{n\to\infty} Y_n(x) \geq y$. Letting $y \uparrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies

$$\liminf_{n\to\infty} Y_n(x) \geq Y(x).$$

Similarly, for $x \notin E$ and $y \in \mathcal{C}(F_0)$ with $Y(x) = Y^+(x) < y$, we have $\lim_{n\to\infty} F_n(y) = F_0(y) > x$ and in particular, $F_n(y) > x$ for almost all n. This implies that $Y_n(x) \leq y$ for a.a. n and hence that $\limsup_{n\to\infty} Y_n(x) \leq y$. Letting $y \downarrow Y(x)$ with $y \in \mathcal{C}(F_0)$ then implies

$$\limsup_{n\to\infty} Y_n(x) \leq Y(x).$$

Hence we have shown, for $x \notin E$, that

$$\limsup_{n\to\infty} Y_n(x) \leq Y(x) \leq \liminf_{n\to\infty} Y_n(x)$$

which shows

$$\lim_{n\to\infty} F_n^{\leftarrow}(x) = \lim_{n\to\infty} Y_n(x) = Y(x) = F^{\leftarrow}(x) \text{ for all } x \notin E. \tag{13.12}$$

■

Definition 13.29. *Two random variables, Y and Z, are said to be of the **same type** if there exists constants, $A > 0$ and $B \in \mathbb{R}$ such that*

$$Z \overset{d}{=} AY + B. \tag{13.13}$$

Alternatively put, if $U(y) := P(Y \leq y)$ and $V(y) := P(Z \leq y)$, then U and V should satisfy,

$$U(y) = P(Y \leq y) = P(Z \leq Ay + B) = V(Ay + B).$$

For the next theorem we will need the following elementary observation.

Lemma 13.30. *If Y is non-constant (a.s.) random variable and $U(y) := P(Y \leq y)$, then $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$ for all γ_1 sufficiently close to 0 and γ_2 sufficiently close to 1.*

Proof. Observe that Y is constant iff $U(y) = 1_{y \geq c}$ for some $c \in \mathbb{R}$, i.e. iff U only takes on the values, $\{0, 1\}$. So since Y is not constant, there exists $y \in \mathbb{R}$ such that $0 < U(y) < 1$. Hence if $\gamma_2 > U(y)$ then $U^{\leftarrow}(\gamma_2) \geq y$ and if $\gamma_1 < U(y)$ then $U^{\leftarrow}(\gamma_1) \leq y$. Moreover, if we suppose that γ_1 is not the height of a flat spot of U, then in fact, $U^{\leftarrow}(\gamma_1) < U^{\leftarrow}(\gamma_2)$. This inequality then remains valid as γ_1 decreases and γ_2 increases. ■

Theorem 13.31 (Convergence of Types). *Suppose* $\{X_n\}_{n=1}^{\infty}$ *is a sequence of random variables and* $a_n, \alpha_n \in (0, \infty)$, $b_n, \beta_n \in \mathbb{R}$ *are constants and* Y *and* Z *are non-constant random variables. Then*

1. *if*
$$\frac{X_n - b_n}{a_n} \implies Y \tag{13.14}$$
and
$$\frac{X_n - \beta_n}{\alpha_n} \implies Z, \tag{13.15}$$
then Y *and* Z *are of the same type. Moreover, the limits,*
$$A = \lim_{n\to\infty} \frac{\alpha_n}{a_n} \in (0, \infty) \text{ and } B := \lim_{n\to\infty} \frac{\beta_n - b_n}{a_n} \tag{13.16}$$
exists and $Y \stackrel{d}{=} AZ + B$.
2. *If the relations in Eq. (13.16) hold then either of the convergences in Eqs. (13.14) or (13.15) implies the others with* Z *and* Y *related by Eq. (13.13).*
3. *If there are some constants,* $a_n > 0$ *and* $b_n \in \mathbb{R}$ *and a non-constant random variable* Y, *such that Eq. (13.14) holds, then Eq. (13.15) holds using* α_n *and* β_n *of the form,*
$$\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1) \text{ and } \beta_n := F_n^{\leftarrow}(\gamma_1) \tag{13.17}$$
for some $0 < \gamma_1 < \gamma_2 < 1$. *If the* F_n *are invertible functions, Eq. (13.17) may be written as*
$$F_n(\beta_n) = \gamma_1 \text{ and } F_n(\alpha_n + \beta_n) = \gamma_2. \tag{13.18}$$

Proof. (2) Assume the limits in Eq. (13.16) hold. If Eq. (13.14) is satisfied, then by Slutsky's Theorem 13.20,

$$\begin{aligned}
\frac{X_n - \beta_n}{\alpha_n} &= \frac{X_n - b_n + b_n - \beta_n}{a_n} \frac{a_n}{\alpha_n} \\
&= \frac{X_n - b_n}{a_n} \frac{a_n}{\alpha_n} - \frac{\beta_n - b_n}{a_n} \frac{a_n}{\alpha_n} \\
&\implies A^{-1}(Y - B) =: Z
\end{aligned}$$

Similarly, if Eq. (13.15) is satisfied, then

$$\frac{X_n - b_n}{a_n} = \frac{X_n - \beta_n}{\alpha_n} \frac{\alpha_n}{a_n} + \frac{\beta_n - b_n}{a_n} \implies AZ + B =: Y.$$

(1) If $F_n(y) := P(X_n \le y)$, then

$$P\left(\frac{X_n - b_n}{a_n} \le y\right) = F_n(a_n y + b_n) \text{ and } P\left(\frac{X_n - \beta_n}{\alpha_n} \le y\right) = F_n(\alpha_n y + \beta_n).$$

By assumption we have

$$F_n\left(a_n y+b_n\right) \implies U\left(y\right) \text{ and } F_n\left(\alpha_n y+\beta_n\right) \implies V\left(y\right).$$

If $w:=\sup\left\{y : F_n\left(a_n y+b_n\right)<x\right\}$, then $a_n w+b_n=F_n^{\leftarrow}\left(x\right)$ and hence

$$\sup\left\{y : F_n\left(a_n y+b_n\right)<x\right\}=\frac{F_n^{\leftarrow}\left(x\right)-b_n}{a_n}.$$

Similarly,

$$\sup\left\{y : F_n\left(\alpha_n y+\beta_n\right)<x\right\}=\frac{F_n^{\leftarrow}\left(x\right)-\beta_n}{\alpha_n}.$$

With these identities, it now follows from the proof of Skorohod's Theorem 13.28 (see Eq. (13.12)) that there exists an at most countable subset, Λ, of $(0,1)$ such that,

$$\begin{aligned}\frac{F_n^{\leftarrow}\left(x\right)-b_n}{a_n}&=\sup\left\{y : F_n\left(a_n y+b_n\right)<x\right\}\to U^{\leftarrow}\left(x\right) \text{ and}\\ \frac{F_n^{\leftarrow}\left(x\right)-\beta_n}{\alpha_n}&=\sup\left\{y : F_n\left(\alpha_n y+\beta_n\right)<x\right\}\to V^{\leftarrow}\left(x\right)\end{aligned}$$

for all $x\notin\Lambda$. Since Y and Z are not constants a.s., we can choose, by Lemma 13.30, $\gamma_1<\gamma_2$ not in Λ such that $U^{\leftarrow}\left(\gamma_1\right)<U^{\leftarrow}\left(\gamma_2\right)$ and $V^{\leftarrow}\left(\gamma_1\right)<V^{\leftarrow}\left(\gamma_2\right)$. In particular it follows that

$$\begin{aligned}\frac{F_n^{\leftarrow}\left(\gamma_2\right)-F_n^{\leftarrow}\left(\gamma_1\right)}{a_n}&=\frac{F_n^{\leftarrow}\left(\gamma_2\right)-b_n}{a_n}-\frac{F_n^{\leftarrow}\left(\gamma_1\right)-b_n}{a_n}\\ &\to U^{\leftarrow}\left(\gamma_2\right)-U^{\leftarrow}\left(\gamma_1\right)>0 \qquad (13.19)\end{aligned}$$

and similarly

$$\frac{F_n^{\leftarrow}\left(\gamma_2\right)-F_n^{\leftarrow}\left(\gamma_1\right)}{\alpha_n}\to V^{\leftarrow}\left(\gamma_2\right)-V^{\leftarrow}\left(\gamma_1\right)>0.$$

Taking ratios of the last two displayed equations shows,

$$\frac{\alpha_n}{a_n}\to A:=\frac{U^{\leftarrow}\left(\gamma_2\right)-U^{\leftarrow}\left(\gamma_1\right)}{V^{\leftarrow}\left(\gamma_2\right)-V^{\leftarrow}\left(\gamma_1\right)}\in\left(0,\infty\right).$$

Moreover,

$$\begin{aligned}&\frac{F_n^{\leftarrow}\left(\gamma_1\right)-b_n}{a_n}\to U^{\leftarrow}\left(\gamma_1\right) \text{ and} \qquad (13.20)\\ &\frac{F_n^{\leftarrow}\left(\gamma_1\right)-\beta_n}{a_n}=\frac{F_n^{\leftarrow}\left(\gamma_1\right)-\beta_n}{\alpha_n}\frac{\alpha_n}{a_n}\to AV^{\leftarrow}\left(\gamma_1\right)\end{aligned}$$

and therefore,

$$\frac{\beta_n - b_n}{a_n} = \frac{F_n^{\leftarrow}(\gamma_1) - \beta_n}{a_n} - \frac{F_n^{\leftarrow}(\gamma_1) - b_n}{a_n} \to AV^{\leftarrow}(\gamma_1) - U^{\leftarrow}(\gamma_1) := B.$$

(3) Now suppose that we define $\alpha_n := F_n^{\leftarrow}(\gamma_2) - F_n^{\leftarrow}(\gamma_1)$ and $\beta_n := F_n^{\leftarrow}(\gamma_1)$, then according to Eqs. (13.19) and (13.20)we have

$$\alpha_n/a_n \to U^{\leftarrow}(\gamma_2) - U^{\leftarrow}(\gamma_1) \in (0,1) \text{ and}$$
$$\frac{\beta_n - b_n}{a_n} \to U^{\leftarrow}(\gamma_1) \text{ as } n \to \infty.$$

Thus we may always center and scale the $\{X_n\}$ using α_n and β_n of the form described in Eq. (13.17). ■

13.5 Weak Convergence Examples

Example 13.32. Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. $\exp(\lambda)$ – random variables, i.e. $X_n \geq 0$ a.s. and $P(X_n \geq x) = e^{-\lambda x}$ for all $x \geq 0$. In this case

$$F(x) := P(X_1 \leq x) = 1 - e^{-\lambda(x \vee 0)}$$

Consider $M_n := \max(X_1, \dots, X_n)$. We have, for $x \geq 0$ and $c_n \in (0,\infty)$ that

$$F_n(x) := P(M_n \leq x) = P\left(\cap_{j=1}^n \{X_j \leq x\}\right)$$
$$= \prod_{j=1}^n P(X_j \leq x) = [F(x)]^n = \left(1 - e^{-\lambda x}\right)^n.$$

We now wish to find $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\frac{M_n - b_n}{a_n} \implies Y$.
1. To this end we note that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(M_n \leq a_n x + b_n)$$
$$= F_n(a_n x + b_n) = [F(a_n x + b_n)]^n.$$

If we demand (c.f. Eq. (13.18) above)

$$P\left(\frac{M_n - b_n}{a_n} \leq 0\right) = F_n(b_n) = [F(b_n)]^n \to \gamma_1 \in (0,1),$$

then $b_n \to \infty$ and we find

$$\ln \gamma_1 \sim n \ln F(b_n) = n \ln\left(1 - e^{-\lambda b_n}\right) \sim -ne^{-\lambda b_n}.$$

From this it follows that $b_n \sim \lambda^{-1} \ln n$. Given this, we now try to find a_n by requiring,

$$P\left(\frac{M_n - b_n}{a_n} \leq 1\right) = F_n(a_n + b_n) = [F(a_n + b_n)]^n \to \gamma_2 \in (0,1).$$

However, by what we have done above, this requires $a_n + b_n \sim \lambda^{-1}\ln n$. Hence we may as well take a_n to be constant and for simplicity we take $a_n = 1$.

2. We now compute

$$\lim_{n\to\infty} P\left(M_n - \lambda^{-1}\ln n \le x\right) = \lim_{n\to\infty}\left(1 - e^{-\lambda\left(x+\lambda^{-1}\ln n\right)}\right)^n = \lim_{n\to\infty}\left(1 - \frac{e^{-\lambda x}}{n}\right)^n = \exp\left(-e^{-\lambda x}\right).$$

Notice that $F(x)$ is a distribution function for some random variable, Y, and therefore we have shown

$$M_n - \frac{1}{\lambda}\ln n \implies Y \text{ as } n \to \infty$$

where $P(Y \le x) = \exp\left(-e^{-\lambda x}\right)$.

Example 13.33. For $p \in (0,1)$, let X_p denote the number of trials to get success in a sequence of independent trials with success probability p. Then $P(X_p > n) = (1-p)^n$ and therefore for $x > 0$,

$$P(pX_p > x) = P\left(X_p > \frac{x}{p}\right) = (1-p)^{\left[\frac{x}{p}\right]} = e^{\left[\frac{x}{p}\right]\ln(1-p)} \sim e^{-p\left[\frac{x}{p}\right]} \to e^{-x} \text{ as } p \to 0.$$

Therefore $pX_p \implies T$ where $T \overset{d}{=} \exp(1)$, i.e. $P(T > x) = e^{-x}$ for $x \ge 0$ or alternatively, $P(T \le y) = 1 - e^{-y\vee 0}$.

Remarks on this example. Let us see in a couple of ways where the appropriate centering and scaling of the X_p come from in this example. For this let $q = 1-p$, then $P(X_p = n) = (1-p)^{n-1}p = q^{n-1}p$ for $n \in \mathbb{N}$. Also let

$$F_p(x) = P(X_p \le x) = P(X_p \le [x]) = 1 - q^{[x]}$$

where $[x] := \sum_{n=1}^{\infty} n \cdot 1_{[n,n+1)}$.

Method 1. Our goal is to choose $a_p > 0$ and $b_p \in \mathbb{R}$ such that $\lim_{p\downarrow 0} F_p(a_p x + b_p)$ exists. As above, we first demand (taking $x = 0$) that

$$\lim_{p\downarrow 0} F_p(b_p) = \gamma_1 \in (0,1).$$

Since, $\gamma_1 \sim F_p(b_p) \sim 1 - q^{b_p}$ we require, $q^{b_p} \sim 1 - \gamma_1$ and hence, $c \sim b_p \ln q = b_p \ln(1-p) \sim -b_p p$. This suggests that we take $b_p = 1/p$ say. Having done this, we would like to choose a_p such that

$$F_0(x) := \lim_{p\downarrow 0} F_p(a_p x + b_p) \text{ exists.}$$

Since,

$$F_0(x) \sim F_p(a_p x + b_p) \sim 1 - q^{a_p x + b_p}$$

this requires that

$$(1-p)^{a_p x + b_p} = q^{a_p x + b_p} \sim 1 - F_0(x)$$

and hence that

$$\ln(1 - F_0(x)) = (a_p x + b_p) \ln q \sim (a_p x + b_p)(-p) = -p a_p x - 1.$$

From this (setting $x = 1$) we see that $p a_p \sim c > 0$. Hence we might take $a_p = 1/p$ as well. We then have

$$F_p(a_p x + b_p) = F_p\left(p^{-1} x + p^{-1}\right) = 1 - (1-p)^{\left[p^{-1}(x+1)\right]}$$

which is equal to 0 if $x \le -1$, and for $x > -1$ we find

$$(1-p)^{\left[p^{-1}(x+1)\right]} = \exp\left(\left[p^{-1}(x+1)\right] \ln(1-p)\right) \to \exp(-(x+1)).$$

Hence we have shown,

$$\lim_{p \downarrow 0} F_p(a_p x + b_p) = [1 - \exp(-(x+1))] 1_{x \ge -1}$$

$$\frac{X_p - 1/p}{1/p} = p X_p - 1 \implies T - 1$$

or again that $p X_p \implies T$.

Method 2. (Center and scale using the first moment and the variance of X_p.) The generating function is given by

$$f(z) := \mathbb{E}\left[z^{X_p}\right] = \sum_{n=1}^{\infty} z^n q^{n-1} p = \frac{pz}{1 - qz}.$$

Observe that $f(z)$ is well defined for $|z| < \frac{1}{q}$ and that $f(1) = 1$, reflecting the fact that $P(X_p \in \mathbb{N}) = 1$, i.e. a success must occur almost surely. Moreover, we have

$$f'(z) = \mathbb{E}\left[X_p z^{X_p - 1}\right], \ f''(z) = \mathbb{E}\left[X_p (X_p - 1) z^{X_p - 2}\right], \dots$$
$$f^{(k)}(z) = \mathbb{E}\left[X_p (X_p - 1) \dots (X_p - k + 1) z^{X_p - k}\right]$$

and in particular,

$$\mathbb{E}\left[X_p (X_p - 1) \dots (X_p - k + 1)\right] = f^{(k)}(1) = \left(\frac{d}{dz}\right)^k |_{z=1} \frac{pz}{1 - qz}.$$

Since

$$\frac{d}{dz} \frac{pz}{1 - qz} = \frac{p(1 - qz) + qpz}{(1 - qz)^2} = \frac{p}{(1 - qz)^2}$$

and

$$\frac{d^2}{dz^2}\frac{pz}{1-qz} = 2\frac{pq}{(1-qz)^3}$$

it follows that

$$\begin{aligned}\mu_p := \mathbb{E}X_p = \frac{p}{(1-q)^2} = \frac{1}{p} \text{ and}\\ \mathbb{E}\left[X_p\left(X_p-1\right)\right] = 2\frac{pq}{(1-q)^3} = \frac{2q}{p^2}.\end{aligned}$$

Therefore,

$$\begin{aligned}\sigma_p^2 = \operatorname{Var}(X_p) &= \mathbb{E}X_p^2 - (\mathbb{E}X_p)^2 = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2\\ &= \frac{2q+p-1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.\end{aligned}$$

Thus, if we had used μ_p and σ_p to center and scale X_p we would have considered,

$$\frac{X_p - \frac{1}{p}}{\frac{\sqrt{1-p}}{p}} = \frac{pX_p - 1}{\sqrt{1-p}} \implies T - 1$$

instead.

Theorem 13.34. *Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $P(X_n = \pm 1) = 1/2$ and let $S_n := X_1 + \cdots + X_n$ – the position of a drunk after n steps. Observe that $|S_n|$ is an odd integer if n is odd and an even integer if n is even. Then $\frac{S_m}{\sqrt{m}} \implies N(0,1)$ as $m \to \infty$.*

Proof. (Sketch of the proof.) We start by observing that $S_{2n} = 2k$ iff

$$\begin{aligned}\#\{i \le 2n : X_i = 1\} = n + k \text{ while}\\ \#\{i \le 2n : X_i = -1\} = 2n - (n+k) = n - k\end{aligned}$$

and therefore,

$$P(S_{2n} = 2k) = \binom{2n}{n+k}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n+k)! \cdot (n-k)!}\left(\frac{1}{2}\right)^{2n}.$$

Recall Stirling's formula states,

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \text{ as } n \to \infty$$

and therefore,

$$\begin{aligned}
&P\left(S_{2n}=2k\right) \\
&\sim \frac{(2n)^{2n}\, e^{-2n}\sqrt{4\pi n}}{(n+k)^{n+k}\, e^{-(n+k)}\sqrt{2\pi\,(n+k)}\cdot(n-k)^{n-k}\, e^{-(n-k)}\sqrt{2\pi\,(n-k)}}\left(\frac{1}{2}\right)^{2n} \\
&= \sqrt{\frac{n}{\pi\,(n+k)\,(n-k)}}\left(1+\frac{k}{n}\right)^{-(n+k)}\cdot\left(1-\frac{k}{n}\right)^{-(n-k)} \\
&= \frac{1}{\sqrt{\pi n}}\sqrt{\frac{1}{\left(1+\frac{k}{n}\right)\left(1-\frac{k}{n}\right)}}\left(1-\frac{k^2}{n^2}\right)^{-n}\cdot\left(1+\frac{k}{n}\right)^{-k}\cdot\left(1-\frac{k}{n}\right)^{k} \\
&= \frac{1}{\sqrt{\pi n}}\left(1-\frac{k^2}{n^2}\right)^{-n}\cdot\left(1+\frac{k}{n}\right)^{-k-1/2}\cdot\left(1-\frac{k}{n}\right)^{k-1/2}.
\end{aligned}$$

So if we let $x := 2k/\sqrt{2n}$, i.e. $k = x\sqrt{n/2}$ and $k/n = \frac{x}{\sqrt{2n}}$, we have

$$\begin{aligned}
&P\left(\frac{S_{2n}}{\sqrt{2n}}=x\right) \\
&\sim \frac{1}{\sqrt{\pi n}}\left(1-\frac{x^2}{2n}\right)^{-n}\cdot\left(1+\frac{x}{\sqrt{2n}}\right)^{-x\sqrt{n/2}-1/2}\cdot\left(1-\frac{x}{\sqrt{2n}}\right)^{x\sqrt{n/2}-1/2} \\
&\sim \frac{1}{\sqrt{\pi n}}e^{x^2/2}\cdot e^{\frac{x}{\sqrt{2n}}\left(-x\sqrt{n/2}-1/2\right)}\cdot e^{-\frac{x}{\sqrt{2n}}\left(x\sqrt{n/2}-1/2\right)} \\
&\sim \frac{1}{\sqrt{\pi n}}e^{-x^2/2},
\end{aligned}$$

wherein we have repeatedly used

$$\left(1+a_n\right)^{b_n} = e^{b_n \ln(1+a_n)} \sim e^{b_n a_n} \text{ when } a_n \to 0.$$

We now compute

$$\begin{aligned}
P\left(a \le \frac{S_{2n}}{\sqrt{2n}} \le b\right) &= \sum_{a\le x\le b} P\left(\frac{S_{2n}}{\sqrt{2n}}=x\right) \\
&= \frac{1}{\sqrt{2\pi}}\sum_{a\le x\le b} e^{-x^2/2}\frac{2}{\sqrt{2n}} \qquad (13.21)
\end{aligned}$$

where the sum is over x of the form, $x = \frac{2k}{\sqrt{2n}}$ with $k \in \{0, \pm1, \dots, \pm n\}$. Since $\frac{2}{\sqrt{2n}}$ is the increment of x as k increases by 1, we see the latter expression in Eq. (13.21) is the Riemann sum approximation to

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

This proves $\frac{S_{2n}}{\sqrt{2n}} \implies N(0,1)$. Since

$$\frac{S_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n} + X_{2n+1}}{\sqrt{2n+1}} = \frac{S_{2n}}{\sqrt{2n}} \frac{1}{\sqrt{1+\frac{1}{2n}}} + \frac{X_{2n+1}}{\sqrt{2n+1}},$$

it follows directly (or see Slutsky's Theorem 13.20) that $\frac{S_{2n+1}}{\sqrt{2n+1}} \implies N(0,1)$ as well. ■

Proposition 13.35. *Suppose that* $\{U_n\}_{n=1}^{\infty}$ *are i.i.d. random variables which are uniformly distributed in* $(0,1)$. *Let* $U_{(k,n)}$ *denote the position of the* k^{th} *– largest number from the list,* $\{U_1, U_2, \ldots, U_n\}$. *Further let* $k(n)$ *be chosen so that* $\lim_{n\to\infty} k(n) = \infty$ *while* $\lim_{n\to\infty} \frac{k(n)}{n} = 0$ *and let*

$$X_n := \frac{U_{(k(n),n)} - k(n)/n}{\frac{\sqrt{k(n)}}{n}}.$$

Then $d_{TV}(X_n, N(0,1)) \to 0$ *as* $n \to \infty$.

Proof. (Sketch only. See Resnick, Proposition 8.2.1 for more details.) Observe that, for $x \in (0,1)$, that

$$P\left(U_{(k,n)} \le x\right) = P\left(\sum_{i=1}^{n} X_i \ge k\right) = \sum_{l=k}^{n} \binom{n}{l} x^l (1-x)^{n-l}.$$

From this it follows that $\rho_n(x) := 1_{(0,1)}(x) \frac{d}{dx} P\left(U_{(k,n)} \le x\right)$ is the probability density for $U_{(k,n)}$. It now turns out that $\rho_n(x)$ is a Beta distribution,

$$\rho_n(x) = \binom{n}{k} k \cdot x^{k-1} (1-x)^{n-k}.$$

Giving a direct computation of this result is not so illuminating. So let us go another route. To do this we are going to estimate, $P\left(U_{(k,n)} \in (x, x+\Delta]\right)$, for $\Delta \in (0,1)$. Observe that if $U_{(k,n)} \in (x, x+\Delta]$, then there must be at least one $U_i \in (x, x+\Delta]$, for otherwise, $U_{(k,n)} \le x+\Delta$ would imply $U_{(k,n)} \le x$ as well and hence $U_{(k,n)} \notin (x, x+\Delta]$. Let

$$\Omega_i := \{U_i \in (x, x+\Delta] \text{ and } U_j \notin (x, x+\Delta] \text{ for } j \ne i\}.$$

Since

$$\begin{aligned} P(U_i, U_j \in (x, x+\Delta] \text{ for some } i \ne j \text{ with } i, j \le n) &\le \sum_{i<j\le n} P(U_i, U_j \in (x, x+\Delta]) \\ &\le \frac{n^2 - n}{2} \Delta^2, \end{aligned}$$

we see that

$$P\left(U_{(k,n)} \in (x, x+\Delta]\right) = \sum_{i=1}^{n} P\left(U_{(k,n)} \in (x, x+\Delta], \Omega_i\right) + O\left(\Delta^2\right)$$
$$= nP\left(U_{(k,n)} \in (x, x+\Delta], \Omega_1\right) + O\left(\Delta^2\right).$$

Now on the set, Ω_1; $U_{(k,n)} \in (x, x+\Delta]$ iff there are exactly $k-1$ of $U_2, \ldots, U_n$ in $[0, x]$ and $n-k$ of these in $[x+\Delta, 1]$. This leads to the conclusion that

$$P\left(U_{(k,n)} \in (x, x+\Delta]\right) = n\binom{n-1}{k-1} x^{k-1} \left(1-(x+\Delta)\right)^{n-k} \Delta + O\left(\Delta^2\right)$$

and therefore,

$$\rho_n(x) = \lim_{\Delta \downarrow 0} \frac{P\left(U_{(k,n)} \in (x, x+\Delta]\right)}{\Delta} = \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k}.$$

By Stirling's formula,

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

Since

$$\left(\frac{k-1}{n}\right)^{(k-1/2)} = \left(\frac{k}{n}\right)^{(k-1/2)} \cdot \left(\frac{k-1}{k}\right)^{(k-1/2)}$$
$$= \left(\frac{k}{n}\right)^{(k-1/2)} \cdot \left(1-\frac{1}{k}\right)^{(k-1/2)}$$
$$\sim e^{-1} \left(\frac{k}{n}\right)^{(k-1/2)}$$

we arrive at

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

By the change of variables formula, with

$$x = \frac{u - k(n)/n}{\frac{\sqrt{k(n)}}{n}}$$

on noting the $du = \frac{\sqrt{k(n)}}{n} dx$, $x = -\sqrt{k(n)}$ at $u = 0$, and

$$\begin{aligned} x &= \frac{1 - k(n)/n}{\frac{\sqrt{k(n)}}{n}} = \frac{n - k(n)}{\sqrt{k(n)}} \\ &= \frac{n}{\sqrt{k(n)}} \left(1 - \frac{k(n)}{n}\right) = \sqrt{n} \sqrt{\frac{n}{k(n)}} \left(1 - \frac{k(n)}{n}\right) =: b_n, \end{aligned}$$

$$\begin{aligned} \mathbb{E}[F(X_n)] &= \int_0^1 \rho_n(u) F\left(\frac{u - k(n)/n}{\frac{\sqrt{k(n)}}{n}}\right) du \\ &= \int_{-\sqrt{k(n)}}^{b_n} \frac{\sqrt{k(n)}}{n} \rho_n\left(\frac{\sqrt{k(n)}}{n} x + k(n)/n\right) F(x)\, du. \end{aligned}$$

Using this information, it is then shown in Resnick that

$$\frac{\sqrt{k(n)}}{n} \rho_n\left(\frac{\sqrt{k(n)}}{n} x + k(n)/n\right) \to \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

which upon an application of Scheffé's Lemma 13.3 completes the proof. ■

Remark 13.36. It is possible to understand the normalization constants in the definition of X_n by computing the mean and the variance of $U_{(n,k)}$. After some computations (see Chapter **??**), one arrives at

$$\begin{aligned} \mathbb{E}U_{(k,n)} &= \int_0^1 \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k} x dx \\ &= \frac{k}{n+1} \sim \frac{k}{n}, \\ \mathbb{E}U_{(k,n)}^2 &= \int_0^1 \frac{n!}{(k-1)! \cdot (n-k)!} x^{k-1} (1-x)^{n-k} x^2 dx \\ &= \frac{(k+1)k}{(n+2)(n+1)} \text{ and} \\ \operatorname{Var}\left(U_{(k,n)}\right) &= \frac{(k+1)k}{(n+2)(n+1)} - \frac{k^2}{(n+1)^2} \\ &= \frac{k}{n+1}\left[\frac{k+1}{n+2} - \frac{k}{n+1}\right] \\ &= \frac{k}{n+1}\left[\frac{n-k+1}{(n+2)(n+1)}\right] \sim \frac{k}{n^2}. \end{aligned}$$

13.6 Compactness and Tightness

Suppose that $\Lambda \subset \mathbb{R}$ is a dense set and F and $\tilde{F}$ are two right continuous functions. If $F = \tilde{F}$ on Λ, then $F = \tilde{F}$ on $\mathbb{R}$. Indeed, for $x \in \mathbb{R}$ we have

$$F(x) = \lim_{\Lambda \ni \lambda \downarrow x} F(\lambda) = \lim_{\Lambda \ni \lambda \downarrow x} \tilde{F}(\lambda) = \tilde{F}(x).$$

Lemma 13.37. *If $G : \Lambda \to \mathbb{R}$ is a non-decreasing function, then*

$$F(x) := G_+(x) := \inf \{G(\lambda) : x < \lambda \in \Lambda\} \tag{13.22}$$

is a non-decreasing right continuous function.

Proof. To show F is right continuous, let $x \in \mathbb{R}$ and $\lambda \in \Lambda$ such that $\lambda > x$. Then for any $y \in (x, \lambda)$,

$$F(x) \leq F(y) = G_+(y) \leq G(\lambda)$$

and therefore,

$$F(x) \leq F(x+) := \lim_{y \downarrow x} F(y) \leq G(\lambda).$$

Since $\lambda > x$ with $\lambda \in \Lambda$ is arbitrary, we may conclude, $F(x) \leq F(x+) \leq G_+(x) = F(x)$, i.e. $F(x+) = F(x)$. ■

Proposition 13.38. *Suppose that $\{F_n\}_{n=1}^{\infty}$ is a sequence of distribution functions and $\Lambda \subset \mathbb{R}$ is a dense set such that $G(\lambda) := \lim_{n\to\infty} F_n(\lambda) \in [0,1]$ exists for all $\lambda \in \Lambda$. If, for all $x \in \mathbb{R}$, we define $F = G_+$ as in Eq. (13.22), then $F_n(x) \to F(x)$ for all $x \in \mathcal{C}(F)$. (Note well; as we have already seen, it is possible that $F(\infty) < 1$ and $F(-\infty) > 0$ so that F need not be a distribution function for a measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.)*

Proof. Suppose that $x, y \in \mathbb{R}$ with $x < y$ and and $s, t \in \Lambda$ are chosen so that $x < s < y < t$. Then passing to the limit in the inequality,

$$F_n(s) \leq F_n(y) \leq F_n(t)$$

implies

$$F(x) = G_+(x) \leq G(s) \leq \liminf_{n\to\infty} F_n(y) \leq \limsup_{n\to\infty} F_n(y) \leq G(t).$$

Taking the infinum over $t \in \Lambda \cap (y, \infty)$ and then letting $x \in \mathbb{R}$ tend up to y, we may conclude

$$F(y-) \leq \liminf_{n\to\infty} F_n(y) \leq \limsup_{n\to\infty} F_n(y) \leq F(y) \text{ for all } y \in \mathbb{R}.$$

This completes the proof, since $F(y-) = F(y)$ for $y \in \mathcal{C}(F)$. ■

The next theorem deals with weak convergence of measures on $(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$. So as not have to introduce any new machinery, the reader should identify $\bar{\mathbb{R}}$ with $[-1,1] \subset \mathbb{R}$ via the map,

$$[-1,1] \ni x \to \tan\left(\frac{\pi}{2}x\right) \in \bar{\mathbb{R}}.$$

Hence a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ may be identified with a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which is supported on $[-1,1]$. Using this identification, we see that a $-\infty$ should only be considered a point of continuity of a distribution function, $F : \bar{\mathbb{R}} \to [0,1]$ iff and only if $F(-\infty) = 0$. On the other hand, ∞ is always a point of continuity.

Theorem 13.39 (Helly's Selection Theorem). *Every sequence of probability measures, $\{\mu_n\}_{n=1}^{\infty}$, on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ has a sub-sequence which is weakly convergent to a probability measure, μ_0 on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$.*

Proof. Using the identification described above, rather than viewing μ_n as probability measures on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$, we may view them as probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ which are supported on $[-1,1]$, i.e. $\mu_n([-1,1]) = 1$. As usual, let

$$F_n(x) := \mu_n((-\infty, x]) = \mu_n((-\infty, x] \cap [-1,1]).$$

Since $\{F_n(x)\}_{n=1}^{\infty} \subset [0,1]$ and $[0,1]$ is compact, for each $x \in \mathbb{R}$ we may find a convergence subsequence of $\{F_n(x)\}_{n=1}^{\infty}$. Hence by Cantor's diagonalization argument we may find a subsequence, $\{G_k := F_{n_k}\}_{k=1}^{\infty}$ of the $\{F_n\}_{n=1}^{\infty}$ such that $G(x) := \lim_{k\to\infty} G_k(x)$ exists for all $x \in \Lambda := \mathbb{Q}$.

Letting $F(x) := G(x+)$ as in Eq. (13.22), it follows from Lemma 13.37 and Proposition 13.38 that $G_k = F_{n_k} \implies F_0$. Moreover, since $G_k(x) = 0$ for all $x \in \mathbb{Q} \cap (-\infty, -1)$ and $G_k(x) = 1$ for all $x \in \mathbb{Q} \cap [1, \infty)$. Therefore, $F_0(x) = 1$ for all $x \geq 1$ and $F_0(x) = 0$ for all $x < -1$ and the corresponding measure, μ_0 is supported on $[-1,1]$. Hence μ_0 may now be transferred back to a measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. ■

Example 13.40. Suppose $\delta_{-n} \implies \delta_{-\infty}$ and $\delta_n \implies \delta_{\infty}$ and $\frac{1}{2}(\delta_n + \delta_{-n}) \implies \frac{1}{2}(\delta_{\infty} + \delta_{-\infty})$. This shows that probability may indeed transfer to the points at $\pm\infty$.

The next question we would like to address is when is the limiting measure, μ_0 on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ concentrated on $\mathbb{R}$. The following notion of tightness is the key to answering this question.

Definition 13.41. *A collection of probability measures, Γ, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is **tight** iff for every $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that*

$$\inf_{\mu \in \Gamma} \mu([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon. \tag{13.23}$$

*We further say that a collection of random variables, $\{X_\lambda : \lambda \in \Lambda\}$ is **tight** iff the collection probability measures, $\left\{P \circ X_\lambda^{-1} : \lambda \in \Lambda\right\}$ is tight. Equivalently put, $\{X_\lambda : \lambda \in \Lambda\}$ is tight iff*

$$\lim_{M\to\infty} \sup_{\lambda \in \Lambda} P(|X_\lambda| \geq M) = 0. \tag{13.24}$$

Observe that the definition of uniform integrability (see Definition 11.25) is considerably stronger than the notion of tightness. It is also worth observing that if $\alpha > 0$ and $C := \sup_{\lambda \in \Lambda} \mathbb{E} |X_\lambda|^\alpha < \infty$, then by Chebyschev's inequality,

$$\sup_\lambda P(|X_\lambda| \geq M) \leq \sup_\lambda \left[\frac{1}{M^\alpha} \mathbb{E} |X_\lambda|^\alpha \right] \leq \frac{C}{M^\alpha} \to 0 \text{ as } M \to \infty$$

and therefore $\{X_\lambda : \lambda \in \Lambda\}$ is tight.

Theorem 13.42. *Let $\Gamma := \{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Then Γ is tight, iff every subsequently limit measure, μ_0, on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ is supported on $\mathbb{R}$. In particular if Γ is tight, there is a weakly convergent subsequence of Γ converging to a probability measure on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$.*

Proof. Suppose that $\mu_{n_k} \implies \mu_0$ with μ_0 being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. As usual, let $F_0(x) := \mu_0([-\infty, x])$. If Γ is tight and $\varepsilon > 0$ is given, we may find $M_\varepsilon < \infty$ such that $M_\varepsilon, -M_\varepsilon \in \mathcal{C}(F_0)$ and $\mu_n([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$ for all n. Hence it follows that

$$\mu_0([-M_\varepsilon, M_\varepsilon]) = \lim_{k\to\infty} \mu_{n_k}([-M_\varepsilon, M_\varepsilon]) \geq 1 - \varepsilon$$

and by letting $\varepsilon \downarrow 0$ we conclude that $\mu_0(\mathbb{R}) = \lim_{\varepsilon \downarrow 0} \mu_0([-M_\varepsilon, M_\varepsilon]) = 1$.

Conversely, suppose there is a subsequence $\{\mu_{n_k}\}_{k=1}^\infty$ such that $\mu_{n_k} \implies \mu_0$ with μ_0 being a probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$ such that $\mu_0(\mathbb{R}) < 1$. In this case $\varepsilon_0 := \mu_0(\{-\infty, \infty\}) > 0$ and hence for all $M < \infty$ we have

$$\mu_0([-M, M]) \leq \mu_0\left(\bar{\mathbb{R}}\right) - \mu_0(\{-\infty, \infty\}) = 1 - \varepsilon_0.$$

By choosing M so that $-M$ and M are points of continuity of F_0, it then follows that

$$\lim_{k\to\infty} \mu_{n_k}([-M, M]) = \mu_0([-M, M]) \leq 1 - \varepsilon_0.$$

Therefore,

$$\inf_{n \in \mathbb{N}} \mu_n(([-M, M])) \leq 1 - \varepsilon_0 \text{ for all } M < \infty$$

and $\{\mu_n\}_{n=1}^\infty$ is not tight. ■

13.7 Weak Convergence in Metric Spaces

(This section may be skipped.)

Definition 13.43. *Let X be a metric space. A sequence of probability measures $\{P_n\}_{n=1}^\infty$ is said to converge weakly to a probability P if $\lim_{n\to\infty} P_n(f) = P(f)$ for all for every $f \in BC(X)$. This is actually weak-* convergence when viewing $P_n \in BC(X)^*$.*

For simplicity we will now assume that X is a complete metric space throughout this section.

Proposition 13.44. *The following are equivalent:*

1. $P_n \stackrel{w}{\to} P$ *as* $n \to \infty$, *i.e.* $P_n(f) \to P(f)$ *for all* $f \in\in BC(X)$.
2. $P_n(f) \to P(f)$ *for every* $f \in BC(X)$ *which is uniformly continuous.*
3. $\limsup_{n\to\infty} P_n(F) \le P(F)$ *for all* $F \sqsubset X$.
4. $\liminf_{n\to\infty} P_n(G) \ge P(G)$ *for all* $G \subset_o X$.
5. $\lim_{n\to\infty} P_n(A) = P(A)$ *for all* $A \in \mathcal{B}$ *such that* $P(\mathrm{bd}(A)) = 0$.

Proof. 1. $\implies$ 2. is obvious. For 2. $\implies$ 3., let

$$\varphi(t) := \begin{cases} 1 & \text{if} \quad t \le 0 \\ 1-t & \text{if } 0 \le t \le 1 \\ 0 & \text{if} \quad t \ge 1 \end{cases} \tag{13.25}$$

and let $f_n(x) := \varphi(nd(x,F))$. Then $f_n \in BC(X,[0,1])$ is uniformly continuous, $0 \le 1_F \le f_n$ for all n and $f_n \downarrow 1_F$ as $n \to \infty$. Passing to the limit $n \to \infty$ in the equation

$$0 \le P_n(F) \le P_n(f_m)$$

gives

$$0 \le \limsup_{n\to\infty} P_n(F) \le P(f_m)$$

and then letting $m \to \infty$ in this inequality implies item 3. 3. $\iff$ 4. Assuming item 3., let $F = G^c$, then

$$\begin{aligned} 1 - \liminf_{n\to\infty} P_n(G) &= \limsup_{n\to\infty}(1 - P_n(G)) = \limsup_{n\to\infty} P_n(G^c) \\ &\le P(G^c) = 1 - P(G) \end{aligned}$$

which implies 4. Similarly 4. $\implies$ 3. 3. $\iff$ 5. Recall that $\mathrm{bd}(A) = \bar{A} \setminus A^o$, so if $P(\mathrm{bd}(A)) = 0$ and 3. (and hence also 4. holds) we have

$$\begin{aligned} \limsup_{n\to\infty} P_n(A) &\le \limsup_{n\to\infty} P_n(\bar{A}) \le P(\bar{A}) = P(A) \text{ and} \\ \liminf_{n\to\infty} P_n(A) &\ge \liminf_{n\to\infty} P_n(A^o) \ge P(A^o) = P(A) \end{aligned}$$

from which it follows that $\lim_{n\to\infty} P_n(A) = P(A)$. Conversely, let $F \sqsubset X$ and set $F_\delta := \{x \in X : \rho(x,F) \le \delta\}$. Then

$$\mathrm{bd}(F_\delta) \subset F_\delta \setminus \{x \in X : \rho(x,F) < \delta\} = A_\delta$$

where $A_\delta := \{x \in X : \rho(x,F) = \delta\}$. Since $\{A_\delta\}_{\delta>0}$ are all disjoint, we must have

$$\sum_{\delta>0} P(A_\delta) \le P(X) \le 1$$

and in particular the set $\Lambda := \{\delta > 0 : P(A_\delta) > 0\}$ is at most countable. Let $\delta_n \notin \Lambda$ be chosen so that $\delta_n \downarrow 0$ as $n \to \infty$, then

$$P(F_{\delta_m}) = \lim_{n\to\infty} P_n(F_{\delta_m}) \geq \limsup_{n\to\infty} P_n(F).$$

Let $m \to \infty$ in this equation to conclude $P(F) \geq \limsup_{n\to\infty} P_n(F)$ as desired. To finish the proof we will now show 3. $\Longrightarrow$ 1. By an affine change of variables it suffices to consider $f \in C(X,(0,1))$ in which case we have

$$\sum_{i=1}^{k} \frac{(i-1)}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}} \leq f \leq \sum_{i=1}^{k} \frac{i}{k} 1_{\left\{\frac{(i-1)}{k} \leq f < \frac{i}{k}\right\}}. \tag{13.26}$$

Let $F_i := \left\{\frac{i}{k} \leq f\right\}$ and notice that $F_k = \emptyset$. Then for any probability P,

$$\sum_{i=1}^{k} \frac{(i-1)}{k} \left[P(F_{i-1}) - P(F_i)\right] \leq P(f) \leq \sum_{i=1}^{k} \frac{i}{k} \left[P(F_{i-1}) - P(F_i)\right]. \tag{13.27}$$

Since

$$\begin{aligned}
&\sum_{i=1}^{k} \frac{(i-1)}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k} \frac{(i-1)}{k} P(F_{i-1}) - \sum_{i=1}^{k} \frac{(i-1)}{k} P(F_i) \\
&\quad = \sum_{i=1}^{k-1} \frac{i}{k} P(F_i) - \sum_{i=1}^{k} \frac{i-1}{k} P(F_i) = \frac{1}{k} \sum_{i=1}^{k-1} P(F_i)
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{i=1}^{k} \frac{i}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k} \frac{i-1}{k} \left[P(F_{i-1}) - P(F_i)\right] + \sum_{i=1}^{k} \frac{1}{k} \left[P(F_{i-1}) - P(F_i)\right] \\
&\quad = \sum_{i=1}^{k-1} P(F_i) + \frac{1}{k},
\end{aligned}$$

Eq. (13.27) becomes,

$$\frac{1}{k} \sum_{i=1}^{k-1} P(F_i) \leq P(f) \leq \frac{1}{k} \sum_{i=1}^{k-1} P(F_i) + 1/k.$$

Using this equation with $P = P_n$ and then with $P = P$ we find

$$\limsup_{n\to\infty} P_n(f) \le \limsup_{n\to\infty}\left[\frac{1}{k}\sum_{i=1}^{k-1} P_n(F_i) + 1/k\right]$$
$$\le \frac{1}{k}\sum_{i=1}^{k-1} P(F_i) + 1/k \le P(f) + 1/k.$$

Since k is arbitrary, $\limsup_{n\to\infty} P_n(f) \le P(f)$. Replacing f by $1-f$ in this inequality also gives $\liminf_{n\to\infty} P_n(f) \ge P(f)$ and hence we have shown $\lim_{n\to\infty} P_n(f) = P(f)$ as claimed. ∎

Theorem 13.45 (Skorohod Theorem). *Let (X,d) be a separable metric space and $\{\mu_n\}_{n=0}^\infty$ be probability measures on $(X,\mathcal{B}_X)$ such that $\mu_n \implies \mu_0$ as $n\to\infty$. Then there exists a probability space, $(\Omega,\mathcal{B},P)$ and measurable functions, $Y_n : \Omega \to X$, such that $\mu_n = P\circ Y_n^{-1}$ for all $n\in\mathbb{N}_0 := \mathbb{N}\cup\{0\}$ and $\lim_{n\to\infty} Y_n = Y$ a.s.*

Proof. See Theorem 4.30 on page 79 of Kallenberg [**?**]. ∎

Definition 13.46. *Let X be a topological space. A collection of probability measures Λ on $(X,\mathcal{B}_X)$ is said to be **tight** if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \in \mathcal{B}_X$ such that $P(K_\varepsilon) \ge 1-\varepsilon$ for all $P\in\Lambda$.*

Theorem 13.47. *Suppose X is a separable metrizable space and $\Lambda = \{P_n\}_{n=1}^\infty$ is a tight sequence of probability measures on $\mathcal{B}_X$. Then there exists a subsequence $\{P_{n_k}\}_{k=1}^\infty$ which is weakly convergent to a probability measure P on $\mathcal{B}_X$.*

Proof. First suppose that X is compact. In this case $C(X)$ is a Banach space which is separable by the Stone – Weirstrass theorem, see Exercise **??**. By the Riesz theorem, Corollary **??**, we know that $C(X)^*$ is in one to one correspondence with the complex measures on $(X,\mathcal{B}_X)$. We have also seen that $C(X)^*$ is metrizable and the unit ball in $C(X)^*$ is weak - * compact, see Theorem **??**. Hence there exists a subsequence $\{P_{n_k}\}_{k=1}^\infty$ which is weak -* convergent to a probability measure P on X. Alternatively, use the cantor's diagonalization procedure on a countable dense set $\Gamma \subset C(X)$ so find $\{P_{n_k}\}_{k=1}^\infty$ such that $\Lambda(f) := \lim_{k\to\infty} P_{n_k}(f)$ exists for all $f\in\Gamma$. Then for $g\in C(X)$ and $f\in\Gamma$, we have

$$\begin{aligned}|P_{n_k}(g) - P_{n_l}(g)| &\le |P_{n_k}(g) - P_{n_k}(f)| + |P_{n_k}(f) - P_{n_l}(f)| \\ &\quad + |P_{n_l}(f) - P_{n_l}(g)| \\ &\le 2\|g-f\|_\infty + |P_{n_k}(f) - P_{n_l}(f)|\end{aligned}$$

which shows

$$\limsup_{n\to\infty} |P_{n_k}(g) - P_{n_l}(g)| \le 2\|g-f\|_\infty .$$

Letting $f\in\Lambda$ tend to g in $C(X)$ shows $\limsup_{n\to\infty} |P_{n_k}(g) - P_{n_l}(g)| = 0$ and hence $\Lambda(g) := \lim_{k\to\infty} P_{n_k}(g)$ for all $g\in C(X)$. It is now clear that $\Lambda(g) \ge 0$

for all $g \geq 0$ so that Λ is a positive linear functional on X and thus there is a probability measure P such that $\Lambda(g) = P(g)$.

General case. By Theorem 18.38 we may assume that X is a subset of a compact metric space which we will denote by $\bar{X}$. We now extend P_n to $\bar{X}$ by setting $\bar{P}_n(A) := \bar{P}_n(A \cap X)$ for all $A \in \mathcal{B}_{\bar{X}}$. By what we have just proved, there is a subsequence $\left\{\bar{P}'_k := \bar{P}_{n_k}\right\}_{k=1}^{\infty}$ such that $\bar{P}'_k$ converges weakly to a probability measure $\bar{P}$ on $\bar{X}$. The main thing we now have to prove is that "$\bar{P}(X) = 1$," this is where the tightness assumption is going to be used. Given $\varepsilon > 0$, let $K_\varepsilon \subset X$ be a compact set such that $\bar{P}_n(K_\varepsilon) \geq 1 - \varepsilon$ for all n. Since K_ε is compact in X it is compact in $\bar{X}$ as well and in particular a closed subset of $\bar{X}$. Therefore by Proposition 13.44

$$\bar{P}(K_\varepsilon) \geq \limsup_{k\to\infty} \bar{P}_k^{'}(K_\varepsilon) = 1 - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows with $X_0 := \cup_{n=1}^{\infty} K_{1/n}$ satisfies $\bar{P}(X_0) = 1$. Because $X_0 \in \mathcal{B}_X \cap \mathcal{B}_{\bar{X}}$, we may view $\bar{P}$ as a measure on $\mathcal{B}_X$ by letting $P(A) := \bar{P}(A \cap X_0)$ for all $A \in \mathcal{B}_X$. Given a closed subset $F \subset X$, choose $\tilde{F} \sqsubset \bar{X}$ such that $F = \tilde{F} \cap X$. Then

$$\limsup_{k\to\infty} P'_k(F) = \limsup_{k\to\infty} \bar{P}'_k(\tilde{F}) \leq \bar{P}(\tilde{F}) = \bar{P}(\tilde{F} \cap X_0) = P(F),$$

which shows $P'_k \xrightarrow{w} P$. ■

14

Characteristic Functions (Fourier Transform)

Definition 14.1. *Given a probability measure,* μ *on* $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ *, let*

$$\hat{\mu}(\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\cdot x} d\mu(x)$$

be the ***Fourier transform or characteristic function of*** μ*. If* $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ *is a random vector on some probability space* $(\Omega, \mathcal{B}, P)$ *, then we let* $f(\lambda) := f_X(\lambda) := \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ *. Of course, if* $\mu := P \circ X^{-1}$*, then* $f_X(\lambda) = \hat{\mu}(\lambda)$ *.*

Notation 14.2 *Given a measure* μ *on a measurable space,* $(\Omega, \mathcal{B})$ *and a function,* $f \in L^1(\mu)$ *, we will often write* $\mu(f)$ *for* $\int_\Omega f d\mu$*.*

Definition 14.3. *Let* μ *and* ν *be two probability measure on* $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ *. The* ***convolution*** *of* μ *and* ν*, denoted* $\mu * \nu$*, is the measure,* $P \circ (X+Y)^{-1}$ *where* $\{X, Y\}$ *are two independent random vectors such that* $P \circ X^{-1} = \mu$ *and* $P \circ Y^{-1} = \nu$*.*

Of course we may give a more direct definition of the convolution of μ and ν by observing for $A \in \mathcal{B}_{\mathbb{R}^n}$ that

$$\begin{aligned}
\mu * \nu(A) &= P(X + Y \in A) \\
&= \int_{\mathbb{R}^n} d\mu(x) \int_{\mathbb{R}^n} d\nu(y) 1_A(x+y) && (14.1) \\
&= \int_{\mathbb{R}^n} \nu(A - x) d\mu(x) && (14.2) \\
&= \int_{\mathbb{R}^n} \mu(A - x) d\nu(x). && (14.3)
\end{aligned}$$

Remark 14.4. Suppose that $d\mu(x) = u(x) dx$ where $u(x) \geq 0$ and $\int_{\mathbb{R}^n} u(x) dx = 1$. Then using the translation invariance of Lebesgue measure and Tonelli's theorem, we have

$$\mu * \nu(f) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x+y) u(x) dx d\nu(y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x) u(x-y) dx d\nu(y)$$

from which it follows that

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\, d\nu(y)\right] dx.$$

If we further assume that $d\nu(x) = v(x)\,dx$, then we have

$$d(\mu * \nu)(x) = \left[\int_{\mathbb{R}^n} u(x-y)\, v(y)\, dy\right] dx.$$

To simplify notation we write,

$$u * v(x) = \int_{\mathbb{R}^n} u(x-y)\, v(y)\, dy = \int_{\mathbb{R}^n} v(x-y)\, u(y)\, dy.$$

Example 14.5. Suppose that $n = 1$, $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $d\nu(x) = 1_{[-1,0]}(x)\,dx$ so that $\nu(A) = \mu(-A)$. In this case

$$d(\mu * \nu)(x) = \left(1_{[0,1]} * 1_{[-1,0]}\right)(x)\,dx$$

where

$$\begin{aligned}\left(1_{[0,1]} * 1_{[-1,0]}\right)(x) &= \int_{\mathbb{R}} 1_{[-1,0]}(x-y)\, 1_{[0,1]}(y)\, dy \\ &= \int_{\mathbb{R}} 1_{[0,1]}(y-x)\, 1_{[0,1]}(y)\, dy \\ &= \int_{\mathbb{R}} 1_{[0,1]+x}(y)\, 1_{[0,1]}(y)\, dy \\ &= m([0,1] \cap (x + [0,1])) = (1 - |x|)_+ .\end{aligned}$$

14.1 Basic Properties of the Characteristic Function

Definition 14.6. *A function $f : \mathbb{R}^n \to \mathbb{C}$ is said to be* ***positive definite****, iff $f(-\lambda) = \overline{f(\lambda)}$ for all $\lambda \in \mathbb{R}^n$ and for all $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ the matrix, $\left(\{f(\lambda_j - \lambda_k)\}_{j,.k=1}^m\right)$ is non-negative. More explicitly we require,*

$$\sum_{j,k=1}^{m} f(\lambda_j - \lambda_k)\, \xi_j \bar{\xi}_k \geq 0 \text{ for all } (\xi_1, \dots, \xi_m) \in \mathbb{C}^m.$$

Notation 14.7 *For $l \in \mathbb{N} \cup \{0\}$, let $C^l(\mathbb{R}^n, \mathbb{C})$ denote the vector space of functions, $f : \mathbb{R}^n \to \mathbb{C}$ which are l - time continuously differentiable. More explicitly, if $\partial_j := \frac{\partial}{\partial x_j}$, then $f \in C^l(\mathbb{R}^n, \mathbb{C})$ iff the partial derivatives, $\partial_{j_1} \dots \partial_{j_k} f$, exist and are continuous for $k = 1, 2, \dots, l$ and all $j_1, \dots, j_k \in \{1, 2, \dots, n\}$.*

Proposition 14.8 (Basic Properties of $\hat{\mu}$). *Let μ and ν be two probability measures on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, then;*

1. $\hat{\mu}(0) = 1$, *and* $|\hat{\mu}(\lambda)| \leq 1$ *for all* λ.
2. $\hat{\mu}(\lambda)$ *is continuous.*
3. $\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ *for all* $\lambda \in \mathbb{R}^n$ *and in particular,* $\hat{\mu}$ *is real valued iff* μ *is symmetric, i.e. iff* $\mu(-A) = \mu(A)$ *for all* $A \in \mathcal{B}_{\mathbb{R}^n}$. *(If* $\mu = P \circ X^{-1}$ *for some random vector* X*, then* μ *is symmetric iff* $X \stackrel{d}{=} -X$*.)*
4. $\hat{\mu}$ *is a positive definite function. (For the converse of this result, see Bochner's Theorem 14.41 below.*
5. *If* $\int_{\mathbb{R}^n} \|x\|^l d\mu(x) < \infty$*, then* $\hat{\mu} \in C^l(\mathbb{R}^n, \mathbb{C})$ *and*
$$\partial_{j_1} \dots \partial_{j_m} \hat{\mu}(\lambda) = \int_{\mathbb{R}^n} (ix_{j_1} \dots ix_{j_m}) e^{i\lambda \cdot x} d\mu(x) \text{ for all } m \leq l.$$
6. *If* X *and* Y *are independent random vectors then*
$$f_{X+Y}(\lambda) = f_X(\lambda) f_Y(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
This may be alternatively expressed as
$$\widehat{\mu * \nu}(\lambda) = \hat{\mu}(\lambda) \hat{\nu}(\lambda) \text{ for all } \lambda \in \mathbb{R}^n.$$
7. *If* $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, *and* $X : \Omega \to \mathbb{R}^n$ *is a random vector, then*
$$f_{aX+b}(\lambda) = e^{i\lambda \cdot b} f_X(a\lambda).$$

Proof. The proof of items 1., 2., 6., and 7. are elementary and will be left to the reader. It also easy to see that $\overline{\hat{\mu}(\lambda)} = \hat{\mu}(-\lambda)$ and $\hat{\mu}(\lambda) = \hat{\mu}(-\lambda)$ if μ is symmetric. Therefore if μ is symmetric, then $\hat{\mu}(\lambda)$ is real. Conversely if $\hat{\mu}(\lambda)$ is real then

$$\hat{\mu}(\lambda) = \hat{\mu}(-\lambda) = \int_{\mathbb{R}^n} e^{i\lambda \cdot x} d\nu(x) = \hat{\nu}(\lambda)$$

where $\nu(A) := \mu(-A)$. The uniqueness Proposition 14.10 below then implies $\mu = \nu$, i.e. μ is symmetric. This proves item 3.

Item 5. follows by induction using Corollary 8.38. For item 4. let $m \in \mathbb{N}$, $\{\lambda_j\}_{j=1}^m \subset \mathbb{R}^n$ and $(\xi_1, \dots, \xi_m) \in \mathbb{C}^m$. Then

$$\begin{aligned}\sum_{j,k=1}^m \hat{\mu}(\lambda_j - \lambda_k) \xi_j \bar{\xi}_k &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i(\lambda_j - \lambda_k) \cdot x} \xi_j \bar{\xi}_k d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{j,k=1}^m e^{i\lambda_j \cdot x} \xi_j \overline{e^{i\lambda_k \cdot x} \xi_k} d\mu(x) \\ &= \int_{\mathbb{R}^n} \left| \sum_{j=1}^m e^{i\lambda_j \cdot x} \xi_j \right|^2 d\mu(x) \geq 0.\end{aligned}$$

■

Example 14.9 (Example 14.5 continued.). Let $d\mu(x) = 1_{[0,1]}(x)\,dx$ and $\nu(A) = \mu(-A)$. Then

$$\begin{aligned}
\hat{\mu}(\lambda) &= \int_0^1 e^{i\lambda x} dx = \frac{e^{i\lambda} - 1}{i\lambda}, \\
\hat{\nu}(\lambda) &= \hat{\mu}(-\lambda) = \overline{\hat{\mu}(\lambda)} = \frac{e^{-i\lambda} - 1}{-i\lambda}, \text{ and} \\
\widehat{\mu * \nu}(\lambda) &= \hat{\mu}(\lambda)\,\hat{\nu}(\lambda) = |\hat{\mu}(\lambda)|^2 = \left|\frac{e^{i\lambda} - 1}{i\lambda}\right|^2 = \frac{2}{\lambda^2}\left[1 - \cos\lambda\right].
\end{aligned}$$

According to example 14.5 we also have $d(\mu * \nu)(x) = (1 - |x|)_+\,dx$ and so directly we find

$$\begin{aligned}
\widehat{\mu * \nu}(\lambda) &= \int_{\mathbb{R}} e^{i\lambda x}(1 - |x|)_+\,dx = \int_{\mathbb{R}} \cos(\lambda x)(1 - |x|)_+\,dx \\
&= 2\int_0^1 (1 - x)\cos\lambda x\ dx = 2\int_0^1 (1 - x)\,d\frac{\sin\lambda x}{\lambda} \\
&= -2\int_0^1 d(1 - x)\,\frac{\sin\lambda x}{\lambda} = 2\int_0^1 \frac{\sin\lambda x}{\lambda} dx = 2\frac{-\cos\lambda x}{\lambda^2}\Big|_{x=0}^{x=1} \\
&= 2\frac{1 - \cos\lambda}{\lambda^2}.
\end{aligned}$$

Proposition 14.10 (Injectivity of the Fourier Transform). *If μ and ν are two probability measure on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ such that $\hat{\mu} = \hat{\nu}$, then $\mu = \nu$.*

Proof. Let $\mathbb{H}$ be the subspace of bounded measurable complex functions, $f : \mathbb{R}^n \to \mathbb{C}$, such that $\mu(f) = \nu(f)$. Then $\mathbb{H}$ is closed under bounded convergence and complex conjugation. Suppose that $\Lambda \subset \mathbb{Z}^d$ is a finite set, $L > 0$ and

$$p(x) = \sum_{\lambda \in \Lambda} a_\lambda e^{i\lambda \cdot x/(2\pi L)} \tag{14.4}$$

with $a_\lambda \in \mathbb{C}$. Then by assumption,

$$\mu(p) = \sum_{\lambda \in \Lambda} a_\lambda \hat{\mu}\left(\frac{\lambda}{2\pi L}\right) = \sum_{\lambda \in \Lambda} a_\lambda \hat{\nu}\left(\frac{\lambda}{2\pi L}\right) = \nu(p)$$

so that $p \in \mathbb{H}$. From the Stone-Weirstrass theorem (see Exercise 14.7 below) or the theory of the Fourier series, any $f \in C(\mathbb{R}^n, \mathbb{C})$ which is L – periodic, (i.e. $f(x + Le_i) = f(x)$ for all $x \in \mathbb{R}^d$ and $i = 1, 2, \ldots, n$) may be uniformly approximated by a trigonometric polynomial of the form in Eq. (14.4), see Exercise 14.8 below. Hence it follows from the bounded convergence theorem that $f \in \mathbb{H}$ for all $f \in C(\mathbb{R}^n, \mathbb{C})$ which are L – periodic. Now suppose $f \in C_c(\mathbb{R}^n, \mathbb{C})$. Then for $L > 0$ sufficiently large the function,

$$f_L(x) := \sum_{\lambda \in \mathbb{Z}^n} f(x + L\lambda),$$

is continuous and L periodic and hence $f_L \in \mathbb{H}$. Since $f_L \to f$ boundedly as $L \to \infty$, we may further conclude that $f \in \mathbb{H}$ as well, i.e. $C_c(\mathbb{R}^n, \mathbb{C}) \subset \mathbb{H}$. An application of the multiplicative system Theorem (see either Theorem 9.3 or Theorem 9.14) implies $\mathbb{H}$ contains all bounded $\sigma(C_c(\mathbb{R}^n, \mathbb{R})) = \mathcal{B}_{\mathbb{R}^n}$ – measurable functions and this certainly implies $\mu = \nu$. ■

For the most part we are now going to stick to the one dimensional case, i.e. X will be a random variable and μ will be a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The following Lemma is a special case of item 4. of Proposition 14.8.

Lemma 14.11. *Suppose $n \in \mathbb{N}$ and X is random variables such that $\mathbb{E}[|X|^n] < \infty$. If $\mu = P \circ X^{-1}$ is the distribution of X, then $\hat{\mu}(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ is C^n – differentiable and*

$$\hat{\mu}^{(l)}(\lambda) = \mathbb{E}\left[(iX)^l e^{i\lambda X}\right] = \int_{\mathbb{R}} (ix)^l e^{i\lambda x} d\mu(x) \text{ for } l = 0, 1, 2, \ldots, n.$$

In particular it follows that

$$\mathbb{E}\left[X^l\right] = \frac{\hat{\mu}^{(l)}(0)}{i^l}.$$

The following theorem is a partial converse to this lemma. Hence the combination of Lemma 14.11 and Theorem 14.12 (see also Corollary 14.34 below) shows that there is a correspondence between the number of moments of X and the differentiability of f_X.

Theorem 14.12. *Let X be a random variable, $m \in \{0, 1, 2, \ldots\}$, $f(\lambda) = \mathbb{E}\left[e^{i\lambda X}\right]$. If $f \in C^{2m}(\mathbb{R}, \mathbb{C})$ such that $g := f^{(2m)}$ is differentiable in a neighborhood of 0 and $g''(0) = f^{(2m+2)}(0)$ exists. Then $\mathbb{E}\left[X^{2m+2}\right] < \infty$ and $f \in C^{2m+2}(\mathbb{R}, \mathbb{C})$.*

Proof. This will be proved by induction on m. We start with $m = 0$ in which case we automatically we know by Proposition 14.8 or Lemma 14.11 that $f \in C(\mathbb{R}, \mathbb{C})$). Since

$$u(\lambda) := \operatorname{Re} f(\lambda) = \mathbb{E}[\cos(\lambda X)],$$

it follows that u is an even function of λ and hence $u' = \operatorname{Re} f'$ is an odd function of λ and in particular, $u'(0) = 0$. By the mean value theorem, to each $\lambda > 0$ with λ near 0, there exists $0 < c_\lambda < \lambda$ such that

$$\frac{u(\lambda) - u(0)}{\lambda} = u'(c_\lambda) = u'(c_\lambda) - u'(0).$$

Therefore,

$$\frac{u(0) - u(\lambda)}{\lambda c_\lambda} = -\frac{u'(c_\lambda) - u'(0)}{c_\lambda} \to -u''(0) \text{ as } \lambda \downarrow 0.$$

Since

$$\mathbb{E}\left[\frac{1-\cos(\lambda X)}{\lambda^2}\right] \le \mathbb{E}\left[\frac{1-\cos(\lambda X)}{\lambda c_\lambda}\right] = \frac{u(0)-u(\lambda)}{\lambda c_\lambda}$$

and $\lim_{\lambda\downarrow 0}\frac{1-\cos(\lambda X)}{\lambda^2} = \frac{1}{2}X^2$, we may apply Fatou's lemma to conclude,

$$\frac{1}{2}\mathbb{E}\left[X^2\right] \le \liminf_{\lambda\downarrow 0}\mathbb{E}\left[\frac{1-\cos(\lambda X)}{\lambda^2}\right] \le -u''(0) < \infty.$$

An application of Lemma 14.11 then implies that $f \in C^2(\mathbb{R},\mathbb{C})$.

For the general induction step we assume the truth of the theorem at level m in which case we know by Lemma 14.11 that

$$f^{(2m)}(\lambda) = (-1)^m\,\mathbb{E}\left[X^{2m}e^{i\lambda X}\right] =: (-1)^m\,g(\lambda).$$

By assumption we know that g is differentiable in a neighborhood of 0 and that $g''(0)$ exists. We now proceed exactly as before but now with $u := \operatorname{Re} g$. So for each $\lambda > 0$ near 0, there exists $c_\lambda \in (0,\lambda)$ such that

$$\frac{u(0)-u(\lambda)}{\lambda c_\lambda} \to -u''(0) \text{ as } \lambda \downarrow 0$$

and

$$\mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda^2}\right] \le \mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda c_\lambda}\right] = \frac{u(0)-u(\lambda)}{\lambda c_\lambda}.$$

Another use of Fatou's lemma gives,

$$\frac{1}{2}\mathbb{E}\left[X^{2m+2}\right] = \liminf_{\lambda\downarrow 0}\mathbb{E}\left[X^{2m}\frac{1-\cos(\lambda X)}{\lambda^2}\right] \le -u''(0) < \infty$$

from which Lemma 14.11 may be used to show $f \in C^{2m+2}(\mathbb{R},\mathbb{C})$. This completes the induction argument. ■

14.2 Examples

Example 14.13. If $-\infty < a < b < \infty$ and $d\mu(x) = \frac{1}{b-a}1_{[a,b]}(x)\,dx$ then

$$\hat{\mu}(\lambda) = \frac{1}{b-a}\int_a^b e^{i\lambda x}dx = \frac{e^{i\lambda b}-e^{i\lambda a}}{i\lambda(b-a)}.$$

If $a = -c$ and $b = c$ with $c > 0$, then

$$\hat{\mu}(\lambda) = \frac{\sin\lambda c}{\lambda c}.$$

Observe that

$$\hat{\mu}(\lambda) = 1 - \frac{1}{3!}\lambda^2 c^2 + \dots$$

and therefore, $\hat{\mu}'(0) = 0$ and $\hat{\mu}''(0) = -\frac{1}{3}c^2$ and hence it follows that

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = \frac{1}{3}c^2.$$

Example 14.14. Suppose Z is a Poisson random variable with mean $a > 0$, i.e. $P(Z = n) = e^{-a}\frac{a^n}{n!}$. Then

$$f_Z(\lambda) = \mathbb{E}\left[e^{i\lambda Z}\right] = e^{-a}\sum_{n=0}^{\infty} e^{i\lambda n}\frac{a^n}{n!} = e^{-a}\sum_{n=0}^{\infty}\frac{\left(ae^{i\lambda}\right)^n}{n!} = \exp\left(a\left(e^{i\lambda} - 1\right)\right).$$

Differentiating this result gives,

$$f_Z'(\lambda) = iae^{i\lambda}\exp\left(a\left(e^{i\lambda} - 1\right)\right) \text{ and}$$
$$f_Z''(\lambda) = \left(-a^2 e^{i2\lambda} - ae^{i\lambda}\right)\exp\left(a\left(e^{i\lambda} - 1\right)\right)$$

from which we conclude,

$$\mathbb{E}Z = \frac{1}{i}f_Z'(0) = a \text{ and } \mathbb{E}Z^2 = -f_Z''(0) = a^2 + a.$$

Therefore, $\mathbb{E}Z = a = \mathrm{Var}(Z)$.

Example 14.15. Suppose T is a positive random variable such that $P(T \geq t + s | T \geq s) = P(T \geq t)$ for all $s, t \geq 0$, or equivalently

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

then $P(T \geq t) = e^{-at}$ for some $a > 0$. (Such exponential random variables are often used to model "waiting times.") The distribution function for T is $F_T(t) := P(T \leq t) = 1 - e^{-a(t \vee 0)}$. Since $F_T(t)$ is piecewise differentiable, the law of T, $\mu := P \circ T^{-1}$, has a density,

$$d\mu(t) = F_T'(t)\,dt = ae^{-at}1_{t \geq 0}dt.$$

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 $\mathrm{Var}(T) = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = a^{-2}$.

Proposition 14.16. *If* $d\mu(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, *then* $\hat{\mu}(\lambda) = e^{-\lambda^2/2}$. *In particular we have*

$$\int_{\mathbb{R}} x d\mu(x) = 0 \text{ and } \int_{\mathbb{R}} x^2 d\mu(x) = 1.$$

Proof. Differentiating the formula,

$$\hat{\mu}(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} e^{i\lambda x} dx,$$

for $\hat{\mu}$ with respect to λ and then integrating by parts implies,

$$\begin{aligned} \hat{\mu}'(\lambda) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{-x^2/2} e^{i\lambda x} dx \\ &= \frac{i}{\sqrt{2\pi}} \int_{\mathbb{R}} \left[-\frac{d}{dx} e^{-x^2/2} \right] e^{i\lambda x} dx \\ &= \frac{i}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-x^2/2} \frac{d}{dx} e^{i\lambda x} dx = -\lambda \hat{\mu}(\lambda). \end{aligned}$$

Solving this equation of $\hat{\mu}(\lambda)$ then implies

$$\hat{\mu}(\lambda) = e^{-\lambda^2/2} \hat{\mu}(0) = e^{-\lambda^2/2} \mu(\mathbb{R}) = e^{-\lambda^2/2}.$$

■

Example 14.17. If μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $n \in \mathbb{N}$, then $\hat{\mu}^n$ is the characteristic function of the probability measure, namely the measure

$$\mu^{*n} := \overbrace{\mu * \cdots * \mu}^{n \text{ times}}. \tag{14.5}$$

Alternatively put, if $\{X_k\}_{k=1}^n$ are i.i.d. random variables with $\mu = P \circ X_k^{-1}$, then

$$f_{X_1 + \cdots + X_n}(\lambda) = f_{X_1}^n(\lambda).$$

Example 14.18. Suppose that $\{\mu_n\}_{n=0}^\infty$ are probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $\{p_n\}_{n=0}^\infty \subset [0,1]$ such that $\sum_{n=0}^\infty p_n = 1$. Then $\sum_{n=0}^\infty p_n \hat{\mu}_n$ is the characteristic function of the probability measure,

$$\mu := \sum_{n=0}^\infty p_n \mu_n.$$

Here is a more interesting interpretation of μ. Let $\{X_n\}_{n=0}^\infty \cup \{T\}$ be independent random variables with $P \circ X_n^{-1} = \mu_n$ and $P(T = n) = p_n$ for all $n \in \mathbb{N}_0$. Then $\mu(A) = P(X_T \in A)$, where $X_T(\omega) := X_{T(\omega)}(\omega)$. Indeed,

$$\mu(A) = P(X_T \in A) = \sum_{n=0}^{\infty} P(X_T \in A, T = n) = \sum_{n=0}^{\infty} P(X_n \in A, T = n)$$
$$= \sum_{n=0}^{\infty} P(X_n \in A, T = n) = \sum_{n=0}^{\infty} p_n \mu_n(A).$$

Let us also observe that

$$\hat{\mu}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[e^{i\lambda X_T} : T = n\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[e^{i\lambda X_n} : T = n\right]$$
$$= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{i\lambda X_n}\right] P(T = n) = \sum_{n=0}^{\infty} p_n \hat{\mu}_n(\lambda).$$

Example 14.19. If μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ then $\sum_{n=0}^{\infty} p_n \hat{\mu}^n$ is the characteristic function of a probability measure, ν, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. In this case, $\nu = \sum_{n=0}^{\infty} p_n \mu^{*n}$ where μ^{*n} is defined in Eq. (14.5). As an explicit example, if $a > 0$ and $p_n = \frac{a^n}{n!} e^{-a}$, then

$$\sum_{n=0}^{\infty} p_n \hat{\mu}^n = \sum_{n=0}^{\infty} \frac{a^n}{n!} e^{-a} \hat{\mu}^n = e^{-a} e^{a\hat{\mu}} = e^{a(\hat{\mu}-1)}$$

is the characteristic function of a probability measure. In other words,

$$f_{X_T}(\lambda) = \mathbb{E}\left[e^{i\lambda X_T}\right] = \exp\left(a\left(f_{X_1}(\lambda) - 1\right)\right).$$

14.3 Continuity Theorem

Lemma 14.20 (Tail Estimate). *Let $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}$ be a random variable and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right]$ be its characteristic function. Then for $a > 0$,*

$$P(|X| \geq a) \leq \frac{a}{2} \int_{-2/a}^{2/a} (1 - f_X(\lambda))\, d\lambda = \frac{a}{2} \int_{-2/a}^{2/a} (1 - \operatorname{Re} f_X(\lambda))\, d\lambda \quad (14.6)$$

Proof. Recall that the Fourier transform of the uniform distribution on $[-c, c]$ is $\frac{\sin \lambda c}{\lambda c}$ and hence

$$\frac{1}{2c} \int_{-c}^{c} f_X(\lambda)\, d\lambda = \frac{1}{2c} \int_{-c}^{c} \mathbb{E}\left[e^{i\lambda X}\right] d\lambda = \mathbb{E}\left[\frac{\sin cX}{cX}\right].$$

Therefore,

$$\frac{1}{2c} \int_{-c}^{c} (1 - f_X(\lambda))\, d\lambda = 1 - \mathbb{E}\left[\frac{\sin cX}{cX}\right] = \mathbb{E}[Y_c] \quad (14.7)$$

where

$$Y_c := 1 - \frac{\sin cX}{cX}.$$

Notice that $Y_c \geq 0$ (see Eq. (14.47)) and moreover, $Y_c \geq 1/2$ if $|cX| \geq 2$. Hence we may conclude

$$\mathbb{E}[Y_c] \geq \mathbb{E}[Y_c : |cX| \geq 2] \geq \mathbb{E}\left[\frac{1}{2} : |cX| \geq 2\right] = \frac{1}{2} P(|X| \geq 2/c).$$

Combining this estimate with Eq. (14.7) shows,

$$\frac{1}{2c} \int_{-c}^{c} (1 - f_X(\lambda))\, d\lambda \geq \frac{1}{2} P(|X| \geq 2/c).$$

Taking $a = 2/c$ in this estimate proves Eq. (14.6). ■

Theorem 14.21 (Continuity Theorem). *Suppose that* $\{\mu_n\}_{n=1}^{\infty}$ *is a sequence of probability measure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *and suppose that* $f(\lambda) := \lim_{n\to\infty} \hat{\mu}_n(\lambda)$ *exists for all* $\lambda \in \mathbb{R}$. *If* f *is continuous at* $\lambda = 0$, *then* f *is the characteristic function of a unique probability measure,* μ, *on* $\mathcal{B}_{\mathbb{R}}$ *and* $\mu_n \implies \mu$ *as* $n \to \infty$.

Proof. By the continuity of f at $\lambda = 0$, for ever $\varepsilon > 0$ we may choose a_ε sufficiently large so that

$$\frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \leq \varepsilon/2.$$

According to Lemma 14.20 and the DCT,

$$\begin{aligned} \mu_n(\{x : |x| \geq a_\varepsilon\}) &\leq \frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} \hat{\mu}_n(\lambda))\, d\lambda \\ &\to \frac{1}{2} a_\varepsilon \int_{-2/a_\varepsilon}^{2/a_\varepsilon} (1 - \operatorname{Re} f(\lambda))\, d\lambda \leq \varepsilon/2. \end{aligned}$$

Hence $\mu_n(\{x : |x| \geq a_\varepsilon\}) \leq \varepsilon$ for all sufficiently large n, say $n \geq N$. By increasing a_ε if necessary we can assure that $\mu_n(\{x : |x| \geq a_\varepsilon\}) \leq \varepsilon$ for all n and hence $\Gamma := \{\mu_n\}_{n=1}^{\infty}$ is tight.

By Theorem 13.42, we may find a subsequence, $\{\mu_{n_k}\}_{k=1}^{\infty}$ and a probability measure μ on $\mathcal{B}_{\mathbb{R}}$ such that $\mu_{n_k} \implies \mu$ as $k \to \infty$. Since $x \to e^{i\lambda x}$ is a bounded and continuous function, it follows that

$$\hat{\mu}(\lambda) = \lim_{k\to\infty} \hat{\mu}_{n_k}(\lambda) = f(\lambda) \text{ for all } \lambda \in \mathbb{R},$$

that is f is the characteristic function of a probability measure, μ.

We now claim that $\mu_n \implies \mu$ as $n \to \infty$. If not, we could find a bounded continuous function, g, such that $\lim_{n\to\infty} \mu_n(g) \neq \mu(g)$ or equivalently, there would exists $\varepsilon > 0$ and a subsequence $\{\mu'_k := \mu_{n_k}\}$ such that

$$\left|\mu(g) - \mu'_k(g)\right| \geq \varepsilon \text{ for all } k \in \mathbb{N}.$$

However by Theorem 13.42 again, there is a further subsequence, $\mu''_l = \mu'_{k_l}$ of μ'_k such that $\mu''_l \implies \nu$ for some probability measure ν. Since $\hat{\nu}(\lambda) = \lim_{l\to\infty} \hat{\mu}''_l(\lambda) = f(\lambda) = \hat{\mu}(\lambda)$, it follows that $\mu = \nu$. This leads to a contradiction since,

$$\varepsilon \leq \lim_{l\to\infty} \left|\mu(g) - \mu''_l(g)\right| = \left|\mu(g) - \nu(g)\right| = 0.$$

■

Remark 14.22. One could also use Bochner's Theorem 14.41 to conclude; if $f(\lambda) := \lim_{n\to\infty} \hat{\mu}_n(\lambda)$ is continuous then f is the characteristic function of a probability measure. Indeed, the condition of a function being positive definite is preserved under taking pointwise limits.

Exercise 14.1. Suppose now $X : (\Omega, \mathcal{B}, P) \to \mathbb{R}^d$ is a random vector and $f_X(\lambda) := \mathbb{E}\left[e^{i\lambda\cdot X}\right]$ is its characteristic function. Show for $a > 0$,

$$P\left(|X|_\infty \geq a\right) \leq 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} \left(1 - f_X(\lambda)\right) d\lambda = 2\left(\frac{a}{4}\right)^d \int_{[-2/a,2/a]^d} \left(1 - \operatorname{Re} f_X(\lambda)\right) d\lambda \tag{14.8}$$

where $|X|_\infty = \max_i |X_i|$ and $d\lambda = d\lambda_1, \dots, d\lambda_d$.

Solution to Exercise (14.1). Working as above, we have

$$\left(\frac{1}{2c}\right)^d \int_{[-c,c]^d} \left(1 - e^{i\lambda\cdot X}\right) d\lambda = 1 - \prod_{j=1}^d \frac{\sin cX_j}{cX_j} =: Y_c, \tag{14.9}$$

where as before, $Y_c \geq 0$ and $Y_c \geq 1/2$ if $c\,|X_j| \geq 2$ for some j, i.e. if $c\,|X|_\infty \geq 2$. Therefore taking expectations of Eq. (14.9) implies,

$$\begin{aligned}\left(\frac{1}{2c}\right)^d \int_{[-c,c]^d} \left(1 - f_X(\lambda)\right) d\lambda &= \mathbb{E}\left[Y_c\right] \geq \mathbb{E}\left[Y_c : |X|_\infty \geq 2/c\right] \\ &\geq \mathbb{E}\left[\frac{1}{2} : |X|_\infty \geq 2/c\right] = \frac{1}{2} P\left(|X|_\infty \geq 2/c\right).\end{aligned}$$

Taking $c = 2/a$ in this expression implies Eq. (14.8).

The following lemma will be needed before giving our first applications of the continuity theorem.

Lemma 14.23. *Suppose that* $\{z_n\}_{n=1}^\infty \subset \mathbb{C}$ *satisfies,* $\lim_{n\to\infty} nz_n = \xi \in \mathbb{C}$, *then*

$$\lim_{n\to\infty} \left(1 + z_n\right)^n = e^{\xi}.$$

Proof. Since $nz_n \to \xi$, it follows that $z_n \sim \frac{\xi}{n} \to 0$ as $n \to \infty$ and therefore by Lemma 14.45 below, $(1+z_n) = e^{\ln(1+z_n)}$ and

$$\ln(1+z_n) = z_n + O\left(z_n^2\right) = z_n + O\left(\frac{1}{n^2}\right).$$

Therefore,

$$(1+z_n)^n = \left[e^{\ln(1+z_n)}\right]^n = e^{n\ln(1+z_n)} = e^{n\left(z_n + O\left(\frac{1}{n^2}\right)\right)} \to e^{\xi} \text{ as } n \to \infty.$$

■

Proposition 14.24 (Weak Law of Large Numbers revisited). *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *are i.i.d. integrable random variables. Then* $\frac{S_n}{n} \xrightarrow{P} \mathbb{E}X_1 =: \mu$.

Proof. Let $f(\lambda) := f_{X_1}(\lambda) = \mathbb{E}\left[e^{i\lambda X_1}\right]$. Then by Taylor's theorem, $f(\lambda) = 1 + i\mu\lambda + o(\lambda)$. Since,

$$f_{\frac{S_n}{n}}(\lambda) = \left[f\left(\frac{\lambda}{n}\right)\right]^n = \left[1 + i\mu\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right]^n$$

it follows from Lemma 14.23 that

$$\lim_{n\to\infty} f_{\frac{S_n}{n}}(\lambda) = e^{i\mu\lambda}$$

which is the characteristic function of the constant random variable, μ. By the continuity Theorem 14.21, it follows that $\frac{S_n}{n} \implies \mu$ and since μ is constant we may apply Lemma 13.19 to conclude $\frac{S_n}{n} \xrightarrow{P} \mu$. ■

Theorem 14.25 (The Basic Central Limit Theorem). *Suppose that* $\{X_n\}_{n=1}^{\infty}$ *are i.i.d. square integrable random variables such that* $\mathbb{E}X_1 = 0$ *and* $\mathbb{E}X_1^2 = 1$. *Then* $\frac{S_n}{\sqrt{n}} \implies N(0,1)$.

Proof. By Theorem 14.21 and Proposition 14.16, it suffices to show

$$\lim_{n\to\infty} \mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = e^{-\lambda^2/2} \text{ for all } \lambda \in \mathbb{R}.$$

Letting $f(\lambda) := \mathbb{E}\left[e^{i\lambda X_1}\right]$, we have by Taylor's theorem (see Eq. (14.43) and (14.46)) that

$$f(\lambda) = 1 - \frac{1}{2}(1+\varepsilon(\lambda))\lambda^2 \tag{14.10}$$

where $\varepsilon(\lambda) \to 0$ as $\lambda \to 0$. Therefore,

$$\begin{aligned} f_{\frac{S_n}{\sqrt{n}}}(\lambda) &= \mathbb{E}\left[e^{i\lambda\frac{S_n}{\sqrt{n}}}\right] = \left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n \\ &= \left[1 - \frac{1}{2}\left(1+\varepsilon\left(\frac{\lambda}{\sqrt{n}}\right)\right)\frac{\lambda^2}{n}\right]^n \to e^{-\lambda^2/2}, \end{aligned}$$

wherein we have used Lemma 14.23 with

$$z_n = -\frac{1}{2}\left(1+\varepsilon\left(\frac{\lambda}{\sqrt{n}}\right)\right)\frac{\lambda^2}{n}.$$

Alternative proof. This proof uses Lemma 15.6 below as follows;

$$\begin{aligned}\left|f_{\frac{S_n}{\sqrt{n}}}(\lambda) - e^{-\lambda^2/2}\right| &= \left|\left[f\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n - \left[e^{-\lambda^2/2n}\right]^n\right| \\ &\le n\left|f\left(\frac{\lambda}{\sqrt{n}}\right) - e^{-\lambda^2/2n}\right| \\ &= n\left|1 - \frac{1}{2}\left(1+\varepsilon\left(\frac{\lambda}{\sqrt{n}}\right)\right)\frac{\lambda^2}{n} - \left(1 - \frac{\lambda^2}{2n} + O\left(\frac{1}{n^2}\right)\right)\right| \\ &\to 0 \text{ as } n \to \infty.\end{aligned}$$

■

Corollary 14.26. *If* $\{X_n\}_{n=1}^\infty$ *are i.i.d. square integrable random variables such that* $\mathbb{E}X_1 = 0$ *and* $\mathbb{E}X_1^2 = 1$*, then*

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}} \le y\right) - P\left(N(0,1) \le y\right)\right| \to 0 \text{ as } n \to \infty. \tag{14.11}$$

Proof. This is a direct consequence of Theorem 14.25 and Exercise 13.3. ■

Berry (1941) and Esseñ (1942) showed there exists a constant, $C < \infty$, such that; if $\rho^3 := \mathbb{E}|X_1|^3 < \infty$, then

$$\sup_{\lambda\in\mathbb{R}}\left|P\left(\frac{S_n}{\sqrt{n}} \le y\right) - P\left(N(0,1) \le y\right)\right| \le C\left(\frac{\rho}{\sigma}\right)^3/\sqrt{n}.$$

In particular the rate of convergence is $n^{-1/2}$. The exact value of the best constant C is still unknown but it is known to be less than 1. We will not prove this theorem here. However we will give a related result in Theorem 14.28 below.

Remark 14.27. It is now a reasonable question to ask "why" is the limiting random variable normal in Theorem 14.25. One way to understand this is, if under the assumptions of Theorem 14.25, we know $\frac{S_n}{\sqrt{n}} \implies L$ where L is some random variable with $\mathbb{E}L = 0$ and $\mathbb{E}L^2 = 1$, then

$$\begin{aligned}\frac{S_{2n}}{\sqrt{2n}} &= \frac{1}{\sqrt{2}}\left(\frac{\sum_{k=1,\ k \text{ odd}}^{2n} X_j}{\sqrt{n}} + \frac{\sum_{k=1,\ k \text{ even}}^{2n} X_j}{\sqrt{n}}\right) \\ &\implies \frac{1}{\sqrt{2}}\left(L_1 + L_2\right)\end{aligned} \tag{14.12}$$

where $L_1 \overset{d}{=} L \overset{d}{=} L_2$ and L_1 and L_2 are independent. To rigorously understand this, using characteristic functions we would conclude from Eq. (14.12) that

$$f_{\frac{S_{2n}}{\sqrt{2n}}}(\lambda) = f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right) f_{\frac{S_n}{\sqrt{n}}}\left(\frac{\lambda}{\sqrt{2}}\right).$$

Passing to the limit in this equation then shows, with $f(\lambda) = \lim_{n\to\infty} f_{\frac{S_n}{\sqrt{n}}}(\lambda) = f_L(\lambda)$, that

$$f(\lambda) = \left[f\left(\frac{\lambda}{\sqrt{2}}\right)\right]^2.$$

Iterating this equation then shows

$$f(\lambda) = \left[f\left(\frac{\lambda}{\left(\sqrt{2}\right)^n}\right)\right]^{2^n} = \left[1 - \frac{1}{2}\left(\frac{\lambda}{\left(\sqrt{2}\right)^n}\right)^2\left(1 + \varepsilon\left(\frac{\lambda}{\left(\sqrt{2}\right)^n}\right)\right)\right]^{2^n}.$$

An application of Lemma 14.23 then shows

$$\begin{aligned} f(\lambda) &= \lim_{n\to\infty}\left[1 - \frac{1}{2}\left(\frac{\lambda}{\left(\sqrt{2}\right)^n}\right)^2\left(1 + \varepsilon\left(\frac{\lambda}{\left(\sqrt{2}\right)^n}\right)\right)\right]^{2^n} \\ &= e^{-\frac{1}{2}\lambda^2} = f_{N(0,1)}(\lambda). \end{aligned}$$

That is we must have $L \overset{d}{=} N(0,1)$.

It is interesting to give another proof of the central limit theorem. For this proof we will assume $\{X_n\}_{n=1}^\infty$ has third moments. The only property about normal random variables that we shall use the proof is that if $\{N_n\}_{n=1}^\infty$ are i.i.d. standard normal random variables, then

$$\frac{T_n}{\sqrt{n}} := \frac{N_1 + \cdots + N_n}{\sqrt{n}} \overset{d}{=} N(0,1).$$

Theorem 14.28 (A Non-Characteristic Proof of the CLT). *Suppose that $\{X_n\}_{n=1}^\infty$ are mean zero variance one i.i.d random variables such that $\mathbb{E}|X_1|^3 < \infty$. Then for $f \in C^3(\mathbb{R})$ with $M := \sup_{x\in\mathbb{R}}\left|f^{(3)}(x)\right| < \infty$,*

$$\left|\mathbb{E}f\left(\frac{S_n}{\sqrt{n}}\right) - \mathbb{E}f(N)\right| \le \frac{1}{\sqrt{n}}\frac{M}{3!}\cdot\mathbb{E}\left[|N|^3 + |X_1|^3\right] \tag{14.13}$$

where $S_n := X_1 + \cdots + X_n$ and $N \overset{d}{=} N(0,1)$.

Proof. Let $\left\{\bar{X}_n, N_n\right\}_{n=1}^\infty$ be independent random variables such that $N_n \overset{d}{=} N(0,1)$ and $\bar{X}_n \overset{d}{=} X_1$. To simplify notation, we will denote $\bar{X}_n$ by X_n. Let $T_n := N_1 + \cdots + N_n$ and for $0 \le k \le n$, let

$$V_k := (N_1 + \cdots + N_k + X_{k+1} + \cdots + X_n) / \sqrt{n}$$

with the convention that $V_n = S_n/\sqrt{n}$ and $V_0 = T_n/\sqrt{n}$. Then by a telescoping series argument, it follows that

$$f\left(S_n/\sqrt{n}\right) - f\left(T_n/\sqrt{n}\right) = f(V_n) - f(V_0) = \sum_{k=1}^{n} \left[f(V_k) - f(V_{k-1})\right]. \quad (14.14)$$

We now make use of Taylor's theorem with integral remainder the form,

$$f(x + \Delta) - f(x) = f'(x)\Delta + \frac{1}{2} f''(x)\Delta^2 + r(x, \Delta)\Delta^3 \quad (14.15)$$

where

$$r(x, \Delta) := \frac{1}{2}\int_0^1 f'''(x + t\Delta)(1-t)^2\, dt.$$

Taking Eq. (14.15) with Δ replaced by δ and subtracting the results then implies

$$f(x + \Delta) - f(x + \delta) = f'(x)(\Delta - \delta) + \frac{1}{2} f''(x)\left(\Delta^2 - \delta^2\right) + \rho(x, \Delta), \quad (14.16)$$

where

$$|\rho(x, \Delta)| = \left|r(x, \Delta)\Delta^3 - r(x, \delta)\delta^3\right| \le \frac{M}{3!}\left[|\Delta|^3 + |\delta|^3\right], \quad (14.17)$$

wherein we have used the simple estimate, $|r(x, \Delta)| \le M/3!$.

If we define $U_k := (N_1 + \cdots + N_{k-1} + X_{k+1} + \cdots + X_n)/\sqrt{n}$, then $V_k = U_k + N_k/\sqrt{n}$ and $V_{k-1} = U_k + X_k/\sqrt{n}$. Hence, using Eq. (14.16) with $x = U_k$, $\Delta = N_k/\sqrt{n}$ and $\delta = X_k/\sqrt{n}$, it follows that

$$\begin{aligned} f(V_k) - f(V_{k-1}) &= f\left(U_k + N_k/\sqrt{n}\right) - f\left(U_k + X_k/\sqrt{n}\right) \\ &= \frac{1}{\sqrt{n}} f'(U_k)(N_k - X_k) + \frac{1}{2n} f''(U_k)\left(N_k^2 - X_k^2\right) + R_k \end{aligned} \quad (14.18)$$

where

$$|R_k| = \frac{M}{3! \cdot n^{3/2}}\left[|N_k|^3 + |X_k|^3\right]. \quad (14.19)$$

Taking expectations of Eq. (14.18) using; Eq. (14.19), $\mathbb{E}N_k = 1 = \mathbb{E}X_k$, $\mathbb{E}N_k^2 = 1 = \mathbb{E}X_k^2$ and the fact that U_k is independent of both X_k and N_k, we find

$$\begin{aligned} |\mathbb{E}[f(V_k) - f(V_{k-1})]| = |\mathbb{E}R_k| &\le \frac{M}{3! \cdot n^{3/2}}\mathbb{E}\left[|N_k|^3 + |X_k|^3\right] \\ &\le \frac{M}{3! \cdot n^{3/2}}\mathbb{E}\left[|N_1|^3 + |X_1|^3\right]. \end{aligned}$$

Combining this estimate with Eq. (14.14) shows,

$$\left|\mathbb{E}\left[f\left(S_n/\sqrt{n}\right)-f\left(T_n/\sqrt{n}\right)\right]\right| = \left|\sum_{k=1}^n \mathbb{E}R_k\right| \le \sum_{k=1}^n \mathbb{E}\left|R_k\right|$$
$$\le \frac{1}{\sqrt{n}}\frac{M}{3!}\cdot\mathbb{E}\left[\left|N_1\right|^3+\left|X_1\right|^3\right].$$

This completes the proof of Eq. (14.13) since $\frac{T_n}{\sqrt{n}} \overset{d}{=} N$ because,

$$f_{\frac{T_n}{\sqrt{n}}}(\lambda) = \left[f_N\left(\frac{\lambda}{\sqrt{n}}\right)\right]^n = \exp\left(-\frac{1}{2}n\frac{\lambda^2}{n}\right) = \exp\left(-\lambda^2/2\right) = f_N(\lambda).$$

■

For more in this direction the reader is advised to look up "Stein's method."

14.4 A Fourier Transform Inversion Formula

Proposition 14.10 guarantees the injectivity of the Fourier transform on the space of probability measures. Our next goal is to find an inversion formula for the Fourier transform. To motivate the construction below, let us first recall a few facts about Fourier series. To keep our exposition as simple as possible, we now restrict ourselves to the one dimensional case.

For $L>0$, let $e_n^L(x) := e^{-i\frac{n}{L}x}$ and let

$$(f,g)_L := \frac{1}{2\pi L}\int_{-\pi L}^{\pi L} f(x)\,\bar{g}(x)\,dx$$

for $f,g\in L^2([-\pi L,\pi L],dx)$. Then it is well known (and fairly elementary to prove) that $\left\{e_n^L : n\in\mathbb{Z}\right\}$ is an orthonormal basis for $L^2([-\pi L,\pi L],dx)$. In particular, if $f\in C_c(\mathbb{R})$ with $\mathrm{supp}(f)\subset[-\pi L,\pi L]$, then for $x\in[-\pi L,\pi L]$,

$$f(x) = \sum_{n\in\mathbb{Z}}\left(f,e_n^L\right)_L e_n^L(x) = \frac{1}{2\pi L}\sum_{n\in\mathbb{Z}}\left(\int_{-\pi L}^{\pi L} f(y)\,e^{i\frac{n}{L}y}dy\right)e^{-i\frac{n}{L}x}$$
$$= \frac{1}{2\pi L}\sum_{n\in\mathbb{Z}}\hat{f}\left(\frac{n}{L}\right)e^{-i\frac{n}{L}x} \tag{14.20}$$

where

$$\hat{f}(\lambda) = \int_{-\infty}^{\infty} f(y)\,e^{i\lambda y}dy.$$

Letting $L\to\infty$ in Eq. (14.20) then suggests that

$$\frac{1}{2\pi L}\sum_{n\in\mathbb{Z}}\hat{f}\left(\frac{n}{L}\right)e^{-i\frac{n}{L}x} \to \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\lambda)\,e^{-i\lambda x}d\lambda$$

and we are lead to expect,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\lambda)\, e^{-i\lambda x} d\lambda.$$

Hence if we now think that $f(x)$ is a probability density and let $d\mu(x) := f(x)\,dx$ so that $\hat{\mu}(\lambda) = \hat{f}(\lambda)$, we should expect

$$\begin{aligned}
\mu([a,b]) &= \int_a^b f(x)\,dx = \int_a^b \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\, e^{-i\lambda x} d\lambda\right] dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\left(\int_a^b e^{-i\lambda x} dx\right) d\lambda \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) d\lambda \\
&= \lim_{c\to\infty} \frac{1}{2\pi}\int_{-c}^{c} \hat{\mu}(\lambda)\left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right) d\lambda.
\end{aligned}$$

This should provide some motivation for Theorem 14.30 below. The following lemma is needed in the proof of the inversion Theorem 14.30 below.

Lemma 14.29. *For $c > 0$, let*

$$S(c) := \frac{1}{2\pi}\int_{-c}^{c} \frac{\sin\lambda}{\lambda} d\lambda. \tag{14.21}$$

Then $S(c) \to \pi$ boundedly as $c \to \infty$ and

$$\int_{-c}^{c} \frac{\sin\lambda y}{\lambda} d\lambda = \operatorname{sgn}(y) S(c\,|y|) \;\; \textit{for all } y \in \mathbb{R}. \tag{14.22}$$

where

$$\operatorname{sgn}(y) = \begin{cases} 1 & \textit{if}\; y > 0 \\ -1 & \textit{if}\; y < 0 \\ 0 & \textit{if}\; y = 0 \end{cases}.$$

Proof. The first assertion has already been dealt with in Example 10.12. We will repeat the argument here for the reader's convenience. By symmetry and Fubini's theorem,

$$\begin{aligned}
S(c) &= \frac{1}{\pi}\int_0^c \frac{\sin\lambda}{\lambda} d\lambda = \frac{1}{\pi}\int_0^c \sin\lambda\left(\int_0^\infty e^{-\lambda t} dt\right) d\lambda \\
&= \frac{1}{\pi}\int_0^\infty dt \int_0^c d\lambda \sin\lambda e^{-\lambda t} \\
&= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{1}{1+t^2} e^{-tc}\left[-\cos c - t\sin c\right] dt,
\end{aligned} \tag{14.23}$$

wherein we have used

$$\begin{aligned}\int_0^c d\lambda \sin \lambda e^{-\lambda t} &= \operatorname{Im} \int_0^c d\lambda e^{i\lambda} e^{-\lambda t} = \operatorname{Im} \int_0^c d\lambda e^{(i-t)\lambda} \\ &= \operatorname{Im} \left(\frac{e^{(i-t)c} - 1}{(i-t)} \right) = \frac{1}{1+t^2} \operatorname{Im} \left(\left[e^{(i-t)c} - 1 \right] (-i-t) \right) \\ &= \frac{1}{1+t^2} \left(e^{-tc} \left[-\cos c - t \sin c \right] + 1 \right)\end{aligned}$$

and

$$\frac{1}{\pi} \int_0^\infty \frac{1}{1+t^2} dt = \frac{1}{2}.$$

The the integral in Eq. (14.23) tends to as $c \to \infty$ by the dominated convergence theorem. The second assertion in Eq. (14.22) is a consequence of the change of variables, $z = \lambda y$. ■

Theorem 14.30 (Fourier Inversion Formula). *If μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $-\infty < a < b < \infty$, then*

$$\lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \hat{\mu}(\lambda) \left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda = \mu((a,b)) + \frac{1}{2} \left(\mu(\{a\}) + \mu(\{b\}) \right).$$

Proof. By Fubini's theorem and Lemma 14.29,

$$\begin{aligned}I(c) &:= \int_{-c}^{c} \hat{\mu}(\lambda) \left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda \\ &= \int_{-c}^{c} \left(\int_{\mathbb{R}} e^{i\lambda x} d\mu(x) \right) \left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) d\lambda \\ &= \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda e^{i\lambda x} \left(\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda} \right) \\ &= \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda \left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right).\end{aligned}$$

Since

$$\operatorname{Im} \left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right) = -\left(\frac{\cos(\lambda(a-x)) - \cos(\lambda(b-x))}{\lambda} \right)$$

is an odd function of λ it follows that

$$\begin{aligned}I(c) &= \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda \operatorname{Re} \left(\frac{e^{-i\lambda(a-x)} - e^{-i\lambda(b-x)}}{i\lambda} \right) \\ &= \int_{\mathbb{R}} d\mu(x) \int_{-c}^{c} d\lambda \left(\frac{\sin \lambda(x-a) - \sin \lambda(x-b)}{\lambda} \right) \\ &= 2\pi \int_{\mathbb{R}} d\mu(x) \left[\operatorname{sgn}(x-a) S(c|x-a|) - \operatorname{sgn}(x-b) S(c|x-b|) \right].\end{aligned}$$

Now letting $c \to \infty$ in this expression (using the DCT) shows

$$\begin{aligned}\lim_{c\to\infty} \frac{1}{2\pi} I(c) &= \frac{1}{2}\int_{\mathbb{R}} d\mu(x)\left[\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)\right] \\ &= \frac{1}{2}\int_{\mathbb{R}} d\mu(x)\left[2\cdot 1_{(a,b)}(x) + 1_{\{a\}}(x) + 1_{\{b\}}(x)\right] \\ &= \mu((a,b)) + \frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right].\end{aligned}$$

■

Corollary 14.31. *Suppose that μ is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu} \in L^1(m)$, then $d\mu = \rho dm$ where ρ is a continuous density on $\mathbb{R}$.*

Proof. The function,

$$\rho(x) := \frac{1}{2\pi}\int_{\mathbb{R}} \hat{\mu}(\lambda) e^{-i\lambda x} d\lambda,$$

is continuous by the dominated convergence theorem. Moreover,

$$\begin{aligned}\int_a^b \rho(x)\,dx &= \frac{1}{2\pi}\int_a^b dx \int_{\mathbb{R}} d\lambda \hat{\mu}(\lambda) e^{-i\lambda x} \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} d\lambda \hat{\mu}(\lambda) \int_a^b dx e^{-i\lambda x} \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} d\lambda \hat{\mu}(\lambda)\left[\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right] \\ &= \frac{1}{2\pi}\lim_{c\to\infty}\int_{-c}^{c} \hat{\mu}(\lambda)\left[\frac{e^{-i\lambda a} - e^{-i\lambda b}}{i\lambda}\right] d\lambda \\ &= \mu((a,b)) + \frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right].\end{aligned}$$

Letting $a \uparrow b$ over $a \in \mathbb{R}$ such that $\mu(\{a\}) = 0$ in this identity shows $\mu(\{b\}) = 0$ for all $b \in \mathbb{R}$. Therefore we have shown

$$\mu((a,b]) = \int_a^b \rho(x)\,dx \text{ for all } -\infty < a < b < \infty.$$

Using one of the multiplicative systems theorems, it is now easy to verify that $\mu(A) = \int_A \rho(x)\,dx$ for all $A \in \mathcal{B}_{\mathbb{R}}$ or $\int_{\mathbb{R}} h d\mu = \int_{\mathbb{R}} h\rho d\mu$ for all bounded measurable functions $h : \mathbb{R} \to \mathbb{R}$. This then implies that $\rho \geq 0$, m – a.e., and the $d\mu = \rho dm$. ■

Example 14.32. Recall from Example 14.9 that

$$\int_{\mathbb{R}} e^{i\lambda x}(1 - |x|)_+\,dx = 2\frac{1 - \cos\lambda}{\lambda^2}.$$

Hence it follows[1] from Corollary 14.31 that

$$(1-|x|)_+ = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1-\cos\lambda}{\lambda^2} e^{-i\lambda x} d\lambda. \tag{14.24}$$

Corollary 14.33. *For all random variables, X, we have*

$$\mathbb{E}|X| = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1-\operatorname{Re} f_X(\lambda)}{\lambda^2} d\lambda. \tag{14.25}$$

Proof. Evaluating Eq. (14.24) at $x=0$ implies

$$1 = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos\lambda}{\lambda^2} d\lambda.$$

Making the change of variables, $\lambda \to M\lambda$, in the above integral then shows

$$M = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1-\cos(\lambda M)}{\lambda^2} d\lambda.$$

Now let $M=|X|$ in this expression and then take expectations to find

$$\mathbb{E}|X| = \frac{1}{\pi}\int_{\mathbb{R}} \mathbb{E}\frac{1-\cos\lambda X}{\lambda^2} d\lambda = \frac{1}{\pi}\int_{\mathbb{R}} \frac{1-\operatorname{Re} f_X(\lambda)}{\lambda^2} d\lambda.$$

■

Suppose that we did not know the value of $c := \int_{-\infty}^{\infty} \frac{1-\cos\lambda}{\lambda^2} d\lambda$ is π, we could still proceed as above to learn

$$\mathbb{E}|X| = \frac{1}{c}\int_{\mathbb{R}} \frac{1-\operatorname{Re} f_X(\lambda)}{\lambda^2} d\lambda.$$

We could then evaluate c by making a judicious choice of X. For example if $X \overset{d}{=} N(0,1)$, we would have on one hand

$$\mathbb{E}|X| = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} |x| e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}}.$$

On the other hand, $f_X(\lambda) = e^{-\lambda^2/2}$ and so

$$\begin{aligned}\sqrt{\frac{2}{\pi}} &= -\frac{1}{c}\int_{\mathbb{R}} \left(1-e^{-\lambda^2/2}\right) d\left(\lambda^{-1}\right) = \frac{1}{c}\int_{\mathbb{R}} d\left(1-e^{-\lambda^2/2}\right)\left(\lambda^{-1}\right) \\ &= \frac{1}{c}\int_{\mathbb{R}} e^{-\lambda^2/2} d\lambda = \frac{\sqrt{2\pi}}{c}\end{aligned}$$

from which it follows, again, that $c=\pi$.

[1] This identity could also be verified directly using residue calculus techniques from complex variables.

Corollary 14.34. *Suppose X is a random variable such that $u(\lambda) := f_X(\lambda)$ continuously differentiable for $\lambda \in (-2\varepsilon, 2\varepsilon)$ for some $\varepsilon > 0$. We further assume*

$$\int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda} d\lambda < \infty. \tag{14.26}$$

Then $\mathbb{E}|X| < \infty$ and $f_X \in C^1(\mathbb{R}, \mathbb{C})$. (Since u is even, u' is odd and $u'(0) = 0$. Hence if $u'(\lambda)$ were α – Hölder continuous for some $\alpha > 0$, then Eq. (14.26) would hold.)

Proof. Let $u(\lambda) := \operatorname{Re} f_X(\lambda) = \mathbb{E}[\cos \lambda X]$ and assume that $u \in C^1((-2\varepsilon, 2\varepsilon), \mathbb{C})$. Then according to Eq. (14.25)

$$\pi \cdot \mathbb{E}|X| = \int_{\mathbb{R}} \frac{1 - u(\lambda)}{\lambda^2} d\lambda = \int_{|\lambda| \le \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda + \int_{|\lambda| > \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda.$$

Since $0 \le 1 - u(\lambda) \le 2$ and $2/\lambda^2$ is integrable for $|\lambda| > \varepsilon$, it suffices to show

$$\infty > \int_{|\lambda| \le \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda = \lim_{\delta \downarrow 0} \int_{\delta \le |\lambda| \le \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda.$$

By an integration by parts we find

$$\begin{aligned}
\int_{\delta \le |\lambda| \le \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda &= \int_{\delta \le |\lambda| \le \varepsilon} (1 - u(\lambda))\, d\left(-\lambda^{-1}\right) \\
&= \frac{u(\lambda) - 1}{\lambda} |_\delta^\varepsilon + \frac{u(\lambda) - 1}{\lambda} |_{-\varepsilon}^{-\delta} - \int_{\delta \le |\lambda| \le \varepsilon} \lambda^{-1} u'(\lambda)\, d\lambda \\
&= -\int_{\delta \le |\lambda| \le \varepsilon} \lambda^{-1} u'(\lambda)\, d\lambda + \frac{u(\varepsilon) - 1}{\varepsilon} - \frac{u(-\varepsilon) - 1}{-\varepsilon} \\
&\quad + \frac{u(-\delta) - 1}{-\delta} - \frac{u(\delta) - 1}{\delta}. \\
&\to -\lim_{\delta \downarrow 0} \int_{\delta \le |\lambda| \le \varepsilon} \lambda^{-1} u'(\lambda)\, d\lambda + \frac{u(\varepsilon) + u(-\varepsilon)}{\varepsilon} + u'(0) - u'(0) \\
&\le \int_{|\lambda| \le \varepsilon} \frac{|u'(\lambda)|}{|\lambda|} d\lambda + \frac{u(\varepsilon) + u(-\varepsilon)}{\varepsilon} \\
&= 2 \int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda} d\lambda + \frac{u(\varepsilon) + u(-\varepsilon)}{\varepsilon} < \infty.
\end{aligned}$$

Passing the limit as $\delta \downarrow 0$ using the fact that $u'(\lambda)$ is an odd function, we learn

$$\begin{aligned}
\int_{|\lambda| \le \varepsilon} \frac{1 - u(\lambda)}{\lambda^2} d\lambda &= \lim_{\delta \downarrow 0} \int_{\delta \le |\lambda| \le \varepsilon} \lambda^{-1} u'(\lambda)\, d\lambda + \frac{u(\varepsilon) + u(-\varepsilon)}{\varepsilon} \\
&\le 2 \int_0^\varepsilon \frac{|u'(\lambda)|}{\lambda} d\lambda + \frac{u(\varepsilon) + u(-\varepsilon)}{\varepsilon} < \infty.
\end{aligned}$$

■

14.5 Exercises

Exercise 14.2. For $x, \lambda \in \mathbb{R}$, let

$$\varphi(\lambda, x) := \begin{cases} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} & \text{if } x \neq 0 \\ \\ -\frac{1}{2}\lambda^2 & \text{if } x = 0. \end{cases}$$

(It is easy to see that $\varphi(\lambda, 0) = \lim_{x \to 0} \varphi(\lambda, x)$ and in fact that $\varphi(\lambda, x)$ is smooth in (λ, x).) Let $\{x_k\}_{k=1}^n \subset \mathbb{R} \setminus \{0\}$, $\{Z_k\}_{k=1}^n \cup \{N\}$ be independent random variables with $N \stackrel{d}{=} N(0, 1)$ and Z_k being Poisson random variables with mean $a_k > 0$, i.e. $P(Z_k = n) = e^{-a_k} \frac{a_k^n}{n!}$ for $n = 0, 1, 2 \dots$. With $Y := \sum_{k=1}^n x_k (Z_k - a_k) + \alpha N$, show

$$f_Y(\lambda) := \mathbb{E}\left[e^{i\lambda Y}\right] = \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right)$$

where ν is the discrete measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ given by

$$\nu = \alpha^2 \delta_0 + \sum_{k=1}^n a_k x_k^2 \delta_{x_k}. \tag{14.27}$$

Exercise 14.3. To each finite and compactly supported measure, ν, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ show there exists a sequence $\{\nu_n\}_{n=1}^\infty$ of finitely supported finite measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu_n \implies \nu$. Here we say ν is compactly supported if there exists $M < \infty$ such that $\nu(\{x : |x| \geq M\}) = 0$ and we say ν is finitely supported if there exists a finite subset, $\Lambda \subset \mathbb{R}$ such that $\nu(\mathbb{R} \setminus \Lambda) = 0$. Please interpret $\nu_n \implies \nu$ to mean,

$$\int_{\mathbb{R}} f d\nu_n \to \int_{\mathbb{R}} f d\nu \text{ for all } f \in BC(\mathbb{R}).$$

Exercise 14.4. Show that if ν is a finite measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, then

$$f(\lambda) := \exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right) \tag{14.28}$$

is the characteristic function of a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Here is an outline to follow. (You may find the calculus estimates in Section 14.8 to be of help.)

1. Show $f(\lambda)$ is continuous.
2. Now suppose that ν is compactly supported. Show, using Exercises 14.2, 14.3, and the continuity Theorem 14.21 that $\exp\left(\int_{\mathbb{R}} \varphi(\lambda, x)\, d\nu(x)\right)$ is the characteristic function of a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
3. For the general case, approximate ν by a sequence of finite measures with compact support as in item 2.

Exercise 14.5 (Exercise 2.3 in [?]). Let μ be the probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, such that $\mu(\{n\}) = p(n) = c\frac{1}{n^2 \ln|n|} 1_{|n|\geq 2}$ with c chosen so that $\sum_{n\in\mathbb{Z}} p(n) = 1$. Show that $\hat{\mu} \in C^1(\mathbb{R}, \mathbb{C})$ even though $\int_{\mathbb{R}} |x|\, d\mu(x) = \infty$. To do this show,

$$g(t) : \sum_{n\geq 2} \frac{1-\cos nt}{n^2 \ln n}$$

is continuously differentiable.

Exercise 14.6 (Polya's Criterioin [?, Problem 26.3 on p. 305.] and [?, p. 104-107.]). Suppose $\varphi(\lambda)$ is a non-negative symmetric continuous function such that $\varphi(0) = 1$, $\varphi(\lambda)$ is non-increasing and convex for $\lambda \geq 0$. Show $\varphi(\lambda) = \hat{\nu}(\lambda)$ for some probability measure, ν, on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Solution to Exercise (14.6). Because of the continuity theorem and some simple limiting arguments, it suffices to prove the result for a function φ as pictured in Figure 14.1. From Example 14.32, we know that $(1-|\lambda|)_+ = \hat{\mu}(\lambda)$

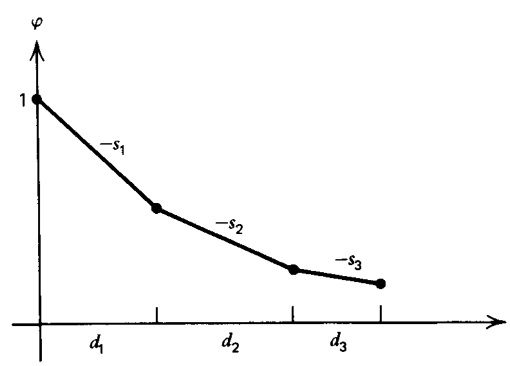


Fig. 14.1. Here is a piecewise linear convex function. We will assume that $d_n > 0$ for all n and that $\varphi(\lambda) = 0$ for λ sufficiently large. This last restriction may be removed later by a limiting argument.

where μ is the probability measure,

$$d\mu(x) := \frac{1}{\pi}\frac{1-\cos x}{x^2} dx.$$

For $a > 0$, let $\mu_a(A) = \mu(aA)$ in which case $\mu_a(f) = \mu\left(f\left(a^{-1}\cdot\right)\right)$ for all bounded measurable f and in particular, $\hat{\mu}_a(\lambda) = \hat{\mu}\left(a^{-1}\lambda\right)$. To finish the proof it suffices to show that $\varphi(\lambda)$ may be expressed as

$$\varphi(\lambda) = \sum_{n=1}^{\infty} p_n \hat{\mu}_{a_n}(\lambda) = \sum_{n=1}^{\infty} p_n \left(1 - \left|\frac{\lambda}{a_n}\right|\right)_+ \qquad (14.29)$$

for some $a_n > 0$ and $p_n \geq 0$ such that $\sum_{n=1}^{\infty} p_n$. Indeed, if this is the case we may take, $\nu := \sum_{n=1}^{\infty} p_n \mu_{a_n}$.

It is pretty clear that we should take $a_n = d_1 + \dots + d_n$ for all $n \in \mathbb{N}$. Since we are assuming $\varphi(\lambda) = 0$ for large λ, there is a first index, $N \in \mathbb{N}$, such that

$$0 = \varphi(a_N) = 1 - \sum_{n=1}^{N} d_n s_n. \tag{14.30}$$

Notice that $s_n = 0$ for all $n > N$.

Since

$$\varphi'(\lambda) = -\sum_{n=k}^{\infty} p_n \frac{1}{a_n} \text{ when } a_{k-1} < \lambda < a_k$$

we must require,

$$s_k = \sum_{n=k}^{\infty} p_n \frac{1}{a_n} \text{ for all } k$$

which then implies $p_k \frac{1}{a_k} = s_k - s_{k+1}$ or equivalently that

$$p_k = a_k (s_k - s_{k+1}). \tag{14.31}$$

Since φ is convex, we know that $-s_k \leq -s_{k+1}$ or $s_k \geq s_{k+1}$ for all k and therefore $p_k \geq 0$ and $p_k = 0$ for all $k > N$. Moreover,

$$\begin{aligned}
\sum_{k=1}^{\infty} p_k &= \sum_{k=1}^{\infty} a_k (s_k - s_{k+1}) = \sum_{k=1}^{\infty} a_k s_k - \sum_{k=2}^{\infty} a_{k-1} s_k \\
&= a_1 s_1 + \sum_{k=2}^{\infty} s_k (a_k - a_{k-1}) = d_1 s_1 + \sum_{k=2}^{\infty} s_k d_k \\
&= \sum_{k=1}^{\infty} s_k d_k = 1
\end{aligned}$$

where the last equality follows from Eq. (14.30). Working backwards with p_k defined as in Eq. (14.31) it is now easily shown that $\frac{d}{d\lambda} \sum_{n=1}^{\infty} p_n \left(1 - \left|\frac{\lambda}{a_n}\right|\right)_+ = \varphi'(\lambda)$ for $\lambda \notin \{a_1, a_2, \dots\}$ and since both functions are equal to 1 at $\lambda = 0$ we may conclude that Eq. (14.29) is indeed valid.

14.6 Appendix: Bochner's Theorem

Definition 14.35. *A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have* ***rapid decay*** *or* ***rapid decrease*** *if*

$$\sup_{x \in \mathbb{R}^n} (1 + |x|)^N |f(x)| < \infty \text{ for } N = 1, 2, \dots.$$

Equivalently, for each $N \in \mathbb{N}$ there exists constants $C_N < \infty$ such that $|f(x)| \leq C_N(1 + |x|)^{-N}$ for all $x \in \mathbb{R}^n$. A function $f \in C(\mathbb{R}^n, \mathbb{C})$ is said to have (at most) ***polynomial growth*** *if there exists $N < \infty$ such*

$$\sup (1 + |x|)^{-N} |f(x)| < \infty,$$

i.e. there exists $N \in \mathbb{N}$ and $C < \infty$ such that $|f(x)| \leq C(1 + |x|)^N$ for all $x \in \mathbb{R}^n$.

Definition 14.36 (Schwartz Test Functions). *Let $\mathcal{S}$ denote the space of functions $f \in C^\infty(\mathbb{R}^n)$ such that f and all of its partial derivatives have rapid decay and let*

$$\|f\|_{N,\alpha} = \sup_{x \in \mathbb{R}^n} \left|(1 + |x|)^N \partial^\alpha f(x)\right|$$

so that

$$\mathcal{S} = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\alpha} < \infty \text{ for all } N \text{ and } \alpha \right\}.$$

Also let $\mathcal{P}$ denote those functions $g \in C^\infty(\mathbb{R}^n)$ such that g and all of its derivatives have at most polynomial growth, i.e. $g \in C^\infty(\mathbb{R}^n)$ is in $\mathcal{P}$ iff for all multi-indices α, there exists $N_\alpha < \infty$ such

$$\sup (1 + |x|)^{-N_\alpha} |\partial^\alpha g(x)| < \infty.$$

(Notice that any polynomial function on $\mathbb{R}^n$ is in $\mathcal{P}$.)

Definition 14.37. *A function $\chi : \mathbb{R}^n \to \mathbb{C}$ is said to be* ***positive (semi) definite*** *iff the matrices $A := \{\chi(\xi_k - \xi_j)\}_{k,j=1}^m$ are positive definite for all $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$.*

Proposition 14.38. *Suppose that $\chi : \mathbb{R}^n \to \mathbb{C}$ is said to be positive definite with $\chi(0) = 1$. If χ is continuous at 0 then in fact χ is uniformly continuous on all of $\mathbb{R}^n$.*

Proof. Taking $\xi_1 = x$, $\xi_2 = y$ and $\xi_3 = 0$ in Definition 14.37 we conclude that

$$A := \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \chi(y-x) & 1 & \chi(y) \\ \chi(-x) & \chi(-y) & 1 \end{bmatrix} = \begin{bmatrix} 1 & \chi(x-y) & \chi(x) \\ \bar{\chi}(x-y) & 1 & \chi(y) \\ \bar{\chi}(x) & \bar{\chi}(y) & 1 \end{bmatrix}$$

is positive definite. In particular,

$$\begin{aligned} 0 \leq \det A &= 1 + \chi(x-y)\chi(y)\bar{\chi}(x) + \chi(x)\bar{\chi}(x-y)\bar{\chi}(y) \\ &\quad - |\chi(x)|^2 - |\chi(y)|^2 - |\chi(x-y)|^2. \end{aligned}$$

Combining this inequality with the identity,

$$|\chi(x) - \chi(y)|^2 = |\chi(x)|^2 + |\chi(y)|^2 - \chi(x)\bar{\chi}(y) - \chi(y)\bar{\chi}(x),$$

gives

$$\begin{aligned}
0 &\le 1 - |\chi(x-y)|^2 + \chi(x-y)\chi(y)\bar\chi(x) + \chi(x)\bar\chi(x-y)\bar\chi(y) \\
&- \left\{|\chi(x) - \chi(y)|^2 + \chi(x)\bar\chi(y) + \chi(y)\bar\chi(x)\right\} \\
&= 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 \\
&+ \chi(x-y)\chi(y)\bar\chi(x) - \chi(y)\bar\chi(x) + \chi(x)\bar\chi(x-y)\bar\chi(y) - \chi(x)\bar\chi(y) \\
&= 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 + 2\,\mathrm{Re}\left((\chi(x-y) - 1)\chi(y)\bar\chi(x)\right) \\
&\le 1 - |\chi(x-y)|^2 - |\chi(x) - \chi(y)|^2 + 2|\chi(x-y) - 1|.
\end{aligned}$$

Hence we have

$$\begin{aligned}
|\chi(x) - \chi(y)|^2 &\le 1 - |\chi(x-y)|^2 + 2|\chi(x-y) - 1| \\
&= (1 - |\chi(x-y)|)(1 + |\chi(x-y)|) + 2|\chi(x-y) - 1| \\
&\le 4|1 - \chi(x-y)|
\end{aligned}$$

which completes the proof. ■

Lemma 14.39. *If $\chi \in C(\mathbb{R}^n, \mathbb{C})$ is a positive definite function, then*

1. $\chi(0) \ge 0$.
2. $\chi(-\xi) = \overline{\chi(\xi)}$ *for all* $\xi \in \mathbb{R}^n$.
3. $|\chi(\xi)| \le \chi(0)$ *for all* $\xi \in \mathbb{R}^n$.
4. *For all* $f \in \mathbb{S}(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \chi(\xi - \eta) f(\xi) \overline{f(\eta)} d\xi d\eta \ge 0. \tag{14.32}$$

Proof. Taking $m = 1$ and $\xi_1 = 0$ we learn $\chi(0)|\lambda|^2 \ge 0$ for all $\lambda \in \mathbb{C}$ which proves item 1. Taking $m = 2$, $\xi_1 = \xi$ and $\xi_2 = \eta$, the matrix

$$A := \begin{bmatrix} \chi(0) & \chi(\xi - \eta) \\ \chi(\eta - \xi) & \chi(0) \end{bmatrix}$$

is positive definite from which we conclude $\chi(\xi - \eta) = \overline{\chi(\eta - \xi)}$ (since $A = A^*$ by definition) and

$$0 \le \det \begin{bmatrix} \chi(0) & \chi(\xi - \eta) \\ \chi(\eta - \xi) & \chi(0) \end{bmatrix} = |\chi(0)|^2 - |\chi(\xi - \eta)|^2.$$

and hence $|\chi(\xi)| \le \chi(0)$ for all ξ. This proves items 2. and 3. Item 4. follows by approximating the integral in Eq. (14.32) by Riemann sums,

$$\begin{aligned}
&\int_{\mathbb{R}^n \times \mathbb{R}^n} \chi(\xi - \eta) f(\xi) \overline{f(\eta)} d\xi d\eta \\
&\qquad = \lim_{\varepsilon \downarrow 0} \varepsilon^{-2n} \sum_{\xi, \eta \in (\varepsilon \mathbb{Z}^n) \cap [-\varepsilon^{-1}, \varepsilon^{-1}]^n} \chi(\xi - \eta) f(\xi) \overline{f(\eta)} \ge 0.
\end{aligned}$$

The details are left to the reader. ■

Lemma 14.40. *If μ is a finite positive measure on $\mathcal{B}_{\mathbb{R}^n}$, then $\chi := \hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$ is a positive definite function.*

Proof. As has already been observed after Definition **??**, the dominated convergence theorem implies $\hat{\mu} \in C(\mathbb{R}^n, \mathbb{C})$. Since μ is a positive measure (and hence real),

$$\hat{\mu}(-\xi) = \int_{\mathbb{R}^n} e^{i\xi\cdot x} d\mu(x) = \overline{\int_{\mathbb{R}^n} e^{-i\xi\cdot x} d\mu(x)} = \overline{\hat{\mu}(-\xi)}.$$

From this it follows that for any $m \in \mathbb{N}$ and $\{\xi_j\}_{j=1}^m \subset \mathbb{R}^n$, the matrix $A := \{\hat{\mu}(\xi_k - \xi_j)\}_{k,j=1}^m$ is self-adjoint. Moreover if $\lambda \in \mathbb{C}^m$,

$$\begin{aligned}\sum_{k,j=1}^m \hat{\mu}(\xi_k - \xi_j)\lambda_k\bar{\lambda}_j &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i(\xi_k-\xi_j)\cdot x}\lambda_k\bar{\lambda}_j d\mu(x) \\ &= \int_{\mathbb{R}^n} \sum_{k,j=1}^m e^{-i\xi_k\cdot x}\lambda_k\overline{e^{-i\xi_j\cdot x}\lambda_j} d\mu(x) \\ &= \int_{\mathbb{R}^n} \left|\sum_{k=1}^m e^{-i\xi_k\cdot x}\lambda_k\right|^2 d\mu(x) \geq 0\end{aligned}$$

showing A is positive definite. ■

Theorem 14.41 (Bochner's Theorem). *Suppose $\chi \in C(\mathbb{R}^n, \mathbb{C})$ is positive definite function, then there exists a unique positive measure μ on $\mathcal{B}_{\mathbb{R}^n}$ such that $\chi = \hat{\mu}$.*

Proof. If $\chi(\xi) = \hat{\mu}(\xi)$, then for $f \in \mathcal{S}$ we would have

$$\int_{\mathbb{R}^n} f d\mu = \int_{\mathbb{R}^n} (f^\vee)\hat{}\ d\mu = \int_{\mathbb{R}^n} f^\vee(\xi)\hat{\mu}(\xi)\mathbf{d}\xi.$$

This suggests that we define

$$I(f) := \int_{\mathbb{R}^n} \chi(\xi)f^\vee(\xi)\mathbf{d}\xi \text{ for all } f \in \mathcal{S}.$$

We will now show I is positive in the sense if $f \in \mathcal{S}$ and $f \geq 0$ then $I(f) \geq 0$. For general $f \in \mathcal{S}$ we have

$$\begin{aligned}I(|f|^2) &= \int_{\mathbb{R}^n} \chi(\xi)\left(|f|^2\right)^\vee(\xi)\mathbf{d}\xi = \int_{\mathbb{R}^n} \chi(\xi)\left(f^\vee \bigstar \bar{f}^\vee\right)(\xi)\mathbf{d}\xi \\ &= \int_{\mathbb{R}^n} \chi(\xi)f^\vee(\xi-\eta)\bar{f}^\vee(\eta)\mathbf{d}\eta\mathbf{d}\xi = \int_{\mathbb{R}^n} \chi(\xi)f^\vee(\xi-\eta)\overline{f^\vee(-\eta)}\mathbf{d}\eta\mathbf{d}\xi \\ &= \int_{\mathbb{R}^n} \chi(\xi-\eta)f^\vee(\xi)\overline{f^\vee(\eta)}\mathbf{d}\eta\mathbf{d}\xi \geq 0. \qquad (14.33)\end{aligned}$$

For $t>0$ let $p_t(x) := t^{-n/2}e^{-|x|^2/2t} \in \mathcal{S}$ and define

$$I_t\left(x\right) := I\bigstar p_t(x) := I(p_t(x-\cdot)) = I(\left|\sqrt{p_t(x-\cdot)}\right|^2)$$

which is non-negative by Eq. (14.33) and the fact that $\sqrt{p_t(x-\cdot)} \in \mathcal{S}$. Using

$$\begin{aligned}\left[p_t(x-\cdot)\right]^\vee(\xi) &= \int_{\mathbb{R}^n} p_t(x-y)e^{iy\cdot\xi}\mathbf{d}y = \int_{\mathbb{R}^n} p_t(y)e^{i(y+x)\cdot\xi}\mathbf{d}y \\ &= e^{ix\cdot\xi}p_t^\vee(\xi) = e^{ix\cdot\xi}e^{-t|\xi|^2/2},\end{aligned}$$

$$\begin{aligned}\langle I_t,\psi\rangle &= \int_{\mathbb{R}^n} I(p_t(x-\cdot))\psi(x)\mathbf{d}x \\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)\left[p_t(x-\cdot)\right]^\vee(\xi)\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}\chi(\xi)e^{ix\cdot\xi}e^{-t|\xi|^2/2}\psi(x)\mathbf{d}\xi\right)\mathbf{d}x \\ &= \int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)e^{-t|\xi|^2/2}\mathbf{d}\xi\end{aligned}$$

which coupled with the dominated convergence theorem shows

$$\langle I\bigstar p_t,\psi\rangle \to \int_{\mathbb{R}^n}\chi(\xi)\psi^\vee(\xi)\mathbf{d}\xi = I(\psi) \text{ as } t\downarrow 0.$$

Hence if $\psi \geq 0$, then $I(\psi) = \lim_{t\downarrow 0}\langle I_t,\psi\rangle \geq 0$.

Let $K \subset \mathbb{R}$ be a compact set and $\psi \in C_c(\mathbb{R},[0,\infty))$ be a function such that $\psi = 1$ on K. If $f \in C_c^\infty(\mathbb{R},\mathbb{R})$ is a smooth function with $\operatorname{supp}(f) \subset K$, then $0 \leq \|f\|_\infty \psi - f \in \mathcal{S}$ and hence

$$0 \leq \langle I, \|f\|_\infty \psi - f\rangle = \|f\|_\infty \langle I,\psi\rangle - \langle I,f\rangle$$

and therefore $\langle I,f\rangle \leq \|f\|_\infty \langle I,\psi\rangle$. Replacing f by $-f$ implies, $-\langle I,f\rangle \leq \|f\|_\infty \langle I,\psi\rangle$ and hence we have proved

$$|\langle I,f\rangle| \leq C(\operatorname{supp}(f))\,\|f\|_\infty \tag{14.34}$$

for all $f \in \mathcal{D}_{\mathbb{R}^n} := C_c^\infty(\mathbb{R}^n,\mathbb{R})$ where $C(K)$ is a finite constant for each compact subset of $\mathbb{R}^n$. Because of the estimate in Eq. (14.34), it follows that $I|_{\mathcal{D}_{\mathbb{R}^n}}$ has a unique extension I to $C_c(\mathbb{R}^n,\mathbb{R})$ still satisfying the estimates in Eq. (14.34) and moreover this extension is still positive. So by the Riesz – Markov Theorem **??**, there exists a unique Radon – measure μ on $\mathbb{R}^n$ such that such that $\langle I,f\rangle = \mu(f)$ for all $f \in C_c(\mathbb{R}^n,\mathbb{R})$.

To finish the proof we must show $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$ given

$$\mu(f) = \int_{\mathbb{R}^n}\chi(\xi)f^\vee(\xi)\mathbf{d}\xi \text{ for all } f \in C_c^\infty(\mathbb{R}^n,\mathbb{R}). \tag{14.35}$$

Let $f \in C_c^\infty(\mathbb{R}^n, \mathbb{R}_+)$ be a radial function such $f(0) = 1$ and $f(x)$ is decreasing as $|x|$ increases. Let $f_\varepsilon(x) := f(\varepsilon x)$, then by Theorem **??**,

$$\mathcal{F}^{-1}\left[e^{-i\eta x} f_\varepsilon(x)\right](\xi) = \varepsilon^{-n} f^\vee\left(\frac{\xi - \eta}{\varepsilon}\right)$$

and therefore, from Eq. (14.35),

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) = \int_{\mathbb{R}^n} \chi(\xi) \varepsilon^{-n} f^\vee\left(\frac{\xi - \eta}{\varepsilon}\right) \mathbf{d}\xi. \tag{14.36}$$

Because $\int_{\mathbb{R}^n} f^\vee(\xi)\mathbf{d}\xi = \mathcal{F} f^\vee(0) = f(0) = 1$, we may apply the approximate δ – function Theorem **??** to Eq. (14.36) to find

$$\int_{\mathbb{R}^n} e^{-i\eta x} f_\varepsilon(x) d\mu(x) \to \chi(\eta) \text{ as } \varepsilon \downarrow 0. \tag{14.37}$$

On the the other hand, when $\eta = 0$, the monotone convergence theorem implies $\mu(f_\varepsilon) \uparrow \mu(1) = \mu(\mathbb{R}^n)$ and therefore $\mu(\mathbb{R}^n) = \mu(1) = \chi(0) < \infty$. Now knowing the μ is a finite measure we may use the dominated convergence theorem to concluded

$$\mu(e^{-i\eta x} f_\varepsilon(x)) \to \mu(e^{-i\eta x}) = \hat{\mu}(\eta) \text{ as } \varepsilon \downarrow 0$$

for all η. Combining this equation with Eq. (14.37) shows $\hat{\mu}(\eta) = \chi(\eta)$ for all $\eta \in \mathbb{R}^n$. ■

14.7 Appendix: A Multi-dimensional Weirstrass Approximation Theorem

The following theorem is the multi-dimensional generalization of Theorem 4.23.

Theorem 14.42 (Weierstrass Approximation Theorem). *Suppose that $K = [a_1, b_1] \times \dots [a_d, b_d]$ with $-\infty < a_i < b_i < \infty$ is a compact rectangle in $\mathbb{R}^d$. Then for every $f \in C(K, \mathbb{C})$, there exists polynomials p_n on $\mathbb{R}^d$ such that $p_n \to f$ uniformly on K.*

Proof. By a simple scaling and translation of the arguments of f we may assume with out loss of generality that $K = [0,1]^d$. By considering the real and imaginary parts of f separately, it suffices to assume $f \in C([\mathbf{0}, \mathbf{1}], \mathbb{R})$.

Given $x \in K$, let $\left\{X_n = \left(X_n^1, \dots, X_n^d\right)\right\}_{n=1}^\infty$ be i.i.d. random vectors with values in $\mathbb{R}^d$ such that

$$P(X_n = \varepsilon) = \prod_{i=1}^d (1 - x_i)^{1-\varepsilon_i} x_i^{\varepsilon_i}$$

for all $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_d) \in \{0,1\}^d$. Since each X_n^j is a Bernoulli random variable with $P\left(X_n^j = 1\right) = x_j$, we know that

$$\mathbb{E}X_n = x \text{ and } \operatorname{Var}\left(X_n^j\right) = x_j - x_j^2 = x_j(1-x_j).$$

As usual let $S_n = S_n := X_1 + \cdots + X_n \in \mathbb{R}^d$, then

$$\mathbb{E}\left[\frac{S_n}{n}\right] = x \text{ and}$$

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{S_n}{n} - x\right\|^2\right] &= \sum_{j=1}^d \mathbb{E}\left(\frac{S_n^j}{n} - x_j\right)^2 = \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n} - x_j\right) \\ &= \sum_{j=1}^d \operatorname{Var}\left(\frac{S_n^j}{n}\right) = \frac{1}{n^2}\cdot\sum_{j=1}^d\sum_{k=1}^n \operatorname{Var}\left(X_k^j\right) \\ &= \frac{1}{n}\sum_{j=1}^d x_j(1-x_j) \le \frac{d}{4n}.\end{aligned}$$

This shows $S_n/n \to x$ in $L^2(P)$ and hence by Chebyshev's inequality, $S_n/n \xrightarrow{P} x$ in and by a continuity theorem, $f\left(\frac{S_n}{n}\right) \xrightarrow{P} f(x)$ as $n \to \infty$. This along with the dominated convergence theorem shows

$$p_n(x) := \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \to f(x) \text{ as } n \to \infty, \tag{14.38}$$

where

$$\begin{aligned}p_n(x) &= \sum_{\varepsilon(\cdot)\in\{0,1\}^d} f\left(\frac{\varepsilon(1) + \cdots + \varepsilon(n)}{n}\right) P\left(X_1 = \varepsilon(1), \ldots, X_n = \varepsilon(n)\right) \\ &= \sum_{\varepsilon(\cdot)\in\{0,1\}^d} f\left(\frac{\varepsilon(1) + \cdots + \varepsilon(n)}{n}\right) \prod_{k=1}^n\prod_{i=1}^d (1-x_i)^{1-\varepsilon_i(k)}\, x_i^{\varepsilon_i(k)}\end{aligned}$$

is a polynomial of degree nd. In fact more is true.

Suppose $\varepsilon > 0$ is given, $M = \sup\{|f(x)| : x \in K\}$, and

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

By uniform continuity of f on K, $\lim_{\varepsilon\downarrow 0}\delta_\varepsilon = 0$. Therefore,

$$\begin{aligned}|f(x) - p_n(x)| &= \left|\mathbb{E}\left(f(x) - f(\frac{S_n}{n})\right)\right| \le \mathbb{E}\left|f(x) - f(\frac{S_n}{n})\right| \\ &\le \mathbb{E}\left[\left|f(x) - f(\frac{S_n}{n})\right| : \|S_n - x\| > \varepsilon\right] \\ &\quad + \mathbb{E}\left[\left|f(x) - f(\frac{S_n}{n})\right| : \|S_n - x\| \le \varepsilon\right] \\ &\le 2MP\left(\|S_n - x\| > \varepsilon\right) + \delta_\varepsilon.\end{aligned} \tag{14.39}$$

By Chebyshev's inequality,

$$P\left(\|S_n - x\| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\mathbb{E}\left\|S_n - x\right\|^2 = \frac{d}{4n\varepsilon^2},$$

and therefore, Eq. (14.39) yields the estimate

$$\sup_{x\in K}\left|f\left(x\right) - p_n\left(x\right)\right| \le \frac{2dM}{n\varepsilon^2} + \delta_\varepsilon$$

and hence

$$\limsup_{n\to\infty}\sup_{x\in K}\left|f\left(x\right) - p_n\left(x\right)\right| \le \delta_\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

■

Here is a version of the complex Weirstrass approximation theorem.

Theorem 14.43 (Complex Weierstrass Approximation Theorem). *Suppose that* $K \subset \mathbb{C}^d \cong \mathbb{R}^d \times \mathbb{R}^d$ *is a compact rectangle. Then there exists polynomials in* $(z = x + iy, \bar{z} = x - iy)$, $p_n(z,\bar{z})$ *for* $z \in \mathbb{C}^d$, *such that* $\sup_{z\in K}\left|q_n(z,\bar{z}) - f(z)\right| \to 0$ *as* $n \to \infty$ *for every* $f \in C\left(K, \mathbb{C}\right)$.

Proof. The mapping $(x,y) \in \mathbb{R}^d\times\mathbb{R}^d \to z = x+iy \in \mathbb{C}^d$ 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 $\mathbb{R}^d \times \mathbb{R}^d$ may be written as a polynomial q in $(z,\bar{z})$, namely

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

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 14.42. ■

Example 14.44. 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 $f \in C\left(S^1\right)$ then $F(z) = |z|\, f(\frac{z}{|z|})$ for $z \neq 0$ and $F(0) = 0$ defines $F \in C(\mathbb{C})$ such that $F|_{S^1} = f$. By applying Theorem 14.43 to F restricted to a compact rectangle containing S^1 we may find $q_n\left(z,\bar{z}\right)$ converging uniformly to F on K and hence on S^1. Since $\bar{z} = z^{-1}$ on S^1, we have shown polynomials in z and z^{-1} are dense in $C(S^1)$. This example generalizes in an obvious way to $K = \left(S^1\right)^d \subset \mathbb{C}^d$.

Exercise 14.7. Use Example 14.44 to show that any 2π – periodic continuous function, $g : \mathbb{R}^d \to \mathbb{C}$, may be uniformly approximated by a trigonometric polynomial of the form

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

where Λ is a finite subset of $\mathbb{Z}^d$ and $a_\lambda \in \mathbb{C}$ for all $\lambda \in \Lambda$. **Hint:** start by showing there exists a unique continuous function, $f : \left(S^1\right)^d \to \mathbb{C}$ such that $f\left(e^{ix_1}, \dots, e^{ix_d}\right) = F\left(x\right)$ for all $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Solution to Exercise (14.7). I will write out the solution when $d = 1$. For $z \in S^1$, define $F(z) := f(e^{i\theta})$ 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. It now follows from Example 14.44 that polynomials in z and z^{-1} are dense in $C(S^1)$. Hence for any $\varepsilon > 0$ there exists

$$p(z, \bar{z}) = \sum a_{m,n} z^m \bar{z}^n = \sum a_{m,n} z^m z^{-n} = \sum a_{m,n} z^{m-n}$$

such that $|F(z) - p(z, \bar{z})| \leq \varepsilon$ for all z. Taking $z = e^{i\theta}$ then implies there exists $b_n \in \mathbb{C}$ and $N \in \mathbb{N}$ such that

$$p_\varepsilon(\theta) := \sum_{n=-N}^{N} b_n e^{in\theta} \tag{14.40}$$

satisfies

$$\sup_\theta \left|\bar{f}(\theta) - p(\theta)\right| \leq \varepsilon.$$

Exercise 14.8. Suppose $f \in C(\mathbb{R}, \mathbb{C})$ is a 2π – periodic function (i.e. $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$) and

$$\int_0^{2\pi} f(x) e^{inx} dx = 0 \text{ for all } n \in \mathbb{Z},$$

show again that $f \equiv 0$. **Hint**: Use Exercise 14.7.

Solution to Exercise (14.8). By assumption, $\int_0^{2\pi} f(\theta) e^{in\theta} d\theta = 0$ for all n and so by the linearity of the Riemann integral,

$$0 = \int_0^{2\pi} f(\theta) p_\varepsilon(\theta) d\theta. \tag{14.41}$$

Choose trigonometric polynomials, p_ε, as in Eq. (14.40) such that $p_\varepsilon(\theta) \to \bar{f}(\theta)$ uniformly in θ as $\varepsilon \downarrow 0$. Passing to the limit in Eq. (14.41) implies

$$0 = \lim_{\varepsilon \downarrow 0} \int_0^{2\pi} f(\theta) p_\varepsilon(\theta) d\theta = \int_0^{2\pi} f(\theta) \bar{f}(\theta) d\theta = \int_0^{2\pi} |f(\theta)|^2 d\theta.$$

From this it follows that $f \equiv 0$, for if $|f(\theta_0)| > 0$ for some θ_0 then $|f(\theta)| \geq \varepsilon > 0$ for θ in a neighborhood of θ_0 by continuity of f. It would then follow that $\int_0^{2\pi} |f(\theta)|^2 d\theta > 0$.

14.8 Appendix: Some Calculus Estimates

We end this section by gathering together a number of calculus estimates that we will need in the future.

1. Taylor's theorem with integral remainder states, if $f \in C^k(\mathbb{R})$ and $z, \Delta \in \mathbb{R}$ or f be holomorphic in a neighborhood of $z \in \mathbb{C}$ and $\Delta \in \mathbb{C}$ be sufficiently small so that $f(z + t\Delta)$ is defined for $t \in [0,1]$, then

$$f(z+\Delta) = \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k r_k(z,\Delta) \tag{14.42}$$

$$= \sum_{n=0}^{k-1} f^{(n)}(z) \frac{\Delta^n}{n!} + \Delta^k \left[\frac{1}{k!} f^{(k)}(z) + \varepsilon(z,\Delta)\right] \tag{14.43}$$

where

$$r_k(z,\Delta) = \frac{1}{(k-1)!} \int_0^1 f^{(k)}(z+t\Delta)(1-t)^{k-1}\, dt \tag{14.44}$$

$$= \frac{1}{k!} f^{(k)}(z) + \varepsilon(z,\Delta) \tag{14.45}$$

and

$$\varepsilon(z,\Delta) = \frac{1}{(k-1)!} \int_0^1 \left[f^{(k)}(z+t\Delta) - f^{(k)}(z)\right](1-t)^{k-1}\, dt \to 0 \text{ as } \Delta \to 0. \tag{14.46}$$

To prove this, use integration by parts to show,

$$\begin{aligned} r_k(z,\Delta) &= \frac{1}{k!} \int_0^1 f^{(k)}(z+t\Delta)\left(-\frac{d}{dt}\right)(1-t)^k\, dt \\ &= -\frac{1}{k!}\left[f^{(k)}(z+t\Delta)(1-t)^k\right]_{t=0}^{t=1} + \frac{\Delta}{k!} \int_0^1 f^{(k+1)}(z+t\Delta)(1-t)^k\, dt \\ &= \frac{1}{k!} f^{(k)}(z) + \Delta r_{k+1}(z,\Delta), \end{aligned}$$

i.e.

$$\Delta^k r_k(z,\Delta) = \frac{1}{k!} f^{(k)}(z)\Delta^k + \Delta^{k+1} r_{k+1}(z,\Delta).$$

The result now follows by induction.

2. For $y \in \mathbb{R}$, $\sin y = y \int_0^1 \cos(ty)\, dt$ and hence

$$|\sin y| \le |y|. \tag{14.47}$$

3. For $y \in \mathbb{R}$ we have

$$\cos y = 1 + y^2 \int_0^1 -\cos(ty)(1-t)\, dt \ge 1 + y^2 \int_0^1 -(1-t)\, dt = 1 - \frac{y^2}{2}.$$

Equivalently put[2],

$$g(y) := \cos y - 1 + y^2/2 \geq 0 \text{ for all } y \in \mathbb{R}. \tag{14.48}$$

4. Since

$$|e^z - 1 - z| = \left| z^2 \int_0^1 e^{tz} (1-t)\, dt \right| \leq |z|^2 \int_0^1 e^{t \operatorname{Re} z} (1-t)\, dt,$$

if $\operatorname{Re} z \leq 0$, then

$$|e^z - 1 - z| \leq |z|^2 /2 \tag{14.49}$$

and if $\operatorname{Re} z > 0$ then

$$|e^z - 1 - z| \leq e^{\operatorname{Re} z} |z|^2 /2.$$

Combining these into one estimate gives,

$$|e^z - 1 - z| \leq e^{0 \vee \operatorname{Re} z} \cdot \frac{|z|^2}{2}. \tag{14.50}$$

5. Since $e^{iy} - 1 = iy \int_0^1 e^{ity} dt$, $\left|e^{iy} - 1\right| \leq |y|$ and hence

$$\left|e^{iy} - 1\right| \leq 2 \wedge |y| \text{ for all } y \in \mathbb{R}. \tag{14.51}$$

Lemma 14.45. *For $z = re^{i\theta}$ with $-\pi < \theta < \pi$ and $r > 0$, let $\ln z = \ln r + i\theta$. Then $\ln : \mathbb{C} \setminus (-\infty, 0] \to \mathbb{C}$ is a holomorphic function such that $e^{\ln z} = z$[3] and if $|z| < 1$ then*

---

[2] Alternatively,

$$|\sin y| = \left| \int_0^y \cos x dx \right| \leq \left| \int_0^y |\cos x|\, dx \right| \leq |y|$$

and for $y \geq 0$ we have,

$$\cos y - 1 = \int_0^y - \sin x dx \geq \int_0^y -x dx = -y^2/2.$$

This last inequality may also be proved as a simple calculus exercise following from; $g(\pm\infty) = \infty$ and $g'(y) = 0$ iff $\sin y = y$ which happens iff $y = 0$.

[3] For the purposes of this lemma it suffices to define $\ln(1+z) = -\sum_{n=1}^\infty (-z)^n /n$ and to then observe: 1)

$$\frac{d}{dz} \ln(1+z) = \sum_{n=0}^\infty (-z)^n = \frac{1}{1+z},$$

and 2) the functions $1+z$ and $e^{\ln(1+z)}$ both solve

$$f'(z) = \frac{1}{1+z} f(z) \text{ with } f(0) = 1$$

and therefore $e^{\ln(1+z)} = 1 + z$.

$$\left|\ln(1+z)-z\right| \le |z|^2 \frac{1}{2(1-|z|)^2} \text{ for } |z|<1. \tag{14.52}$$

Proof. Clearly $e^{\ln z} = z$ and $\ln z$ is continuous. Therefore by the inverse function theorem for holomorphic functions, $\ln z$ is holomorphic and

$$z\frac{d}{dz}\ln z = e^{\ln z}\frac{d}{dz}\ln z = 1.$$

Therefore, $\frac{d}{dz}\ln z = \frac{1}{z}$ and $\frac{d^2}{dz^2}\ln z = -\frac{1}{z^2}$. So by Taylor's theorem,

$$\ln(1+z) = z - z^2\int_0^1 \frac{1}{(1+tz)^2}(1-t)\,dt. \tag{14.53}$$

If $t \ge 0$ and $|z|<1$, then

$$\left|\frac{1}{(1+tz)}\right| \le \sum_{n=0}^\infty |tz|^n = \frac{1}{1-t|z|} \le \frac{1}{1-|z|}.$$

and therefore,

$$\left|\int_0^1 \frac{1}{(1+tz)^2}(1-t)\,dt\right| \le \frac{1}{2(1-|z|)^2}. \tag{14.54}$$

Eq. (14.52) is now a consequence of Eq. (14.53) and Eq. (14.54). ■

Lemma 14.46. *For all $y \in \mathbb{R}$ and $n \in \mathbb{N}\cup\{0\}$,*

$$\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| \le \frac{|y|^{n+1}}{(n+1)!} \tag{14.55}$$

and in particular,

$$\left|e^{iy} - \left(1+iy-\frac{y^2}{2!}\right)\right| \le y^2 \wedge \frac{|y|^3}{3!}. \tag{14.56}$$

More generally for all $n \in \mathbb{N}$ we have

$$\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| \le \frac{|y|^{n+1}}{(n+1)!} \wedge \frac{2|y|^n}{n!}. \tag{14.57}$$

Proof. By Taylor's theorem (see Eq. (14.42) with $f(y) = e^{iy}$, $x = 0$ and $\Delta = y$) we have

$$\begin{aligned}\left|e^{iy} - \sum_{k=0}^n \frac{(iy)^k}{k!}\right| &= \left|\frac{y^{n+1}}{n!}\int_0^1 i^{n+1}e^{ity}(1-t)^n\,dt\right| \\ &\le \frac{|y|^{n+1}}{n!}\int_0^1 (1-t)^n\,dt = \frac{|y|^{n+1}}{(n+1)!}\end{aligned}$$

which is Eq. (14.55). Using Eq. (14.55) with $n = 1$ implies

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \le \left|e^{iy} - (1 + iy)\right| + \left|\frac{y^2}{2}\right| \le \left|\frac{y^2}{2}\right| + \left|\frac{y^2}{2}\right| = y^2$$

and using Eq. (14.55) with $n = 2$ implies

$$\left|e^{iy} - \left(1 + iy - \frac{y^2}{2!}\right)\right| \le \frac{|y|^3}{3!}.$$

Combining the last two inequalities completes the proof of Eq. (14.56). Equation (14.57) is proved similarly and hence will be omitted. ∎

Lemma 14.47. *If X is a square integrable random variable, then*

$$f(\lambda) := \mathbb{E}\left[e^{i\lambda X}\right] = 1 + i\lambda \mathbb{E}X - \frac{\lambda^2}{2!}\mathbb{E}\left[X^2\right] + r(\lambda)$$

where

$$r(\lambda) := \lambda^2 \mathbb{E}\left[X^2 \wedge \frac{|\lambda|\,|X|^3}{3!}\right] = \lambda^2 \varepsilon(\lambda)$$

and

$$\varepsilon(\lambda) := \mathbb{E}\left[X^2 \wedge \frac{|\lambda|\,|X|^3}{3!}\right] \to 0 \text{ as } \lambda \to 0. \tag{14.58}$$

Proof. Using Eq. (14.56) with $y = \lambda X$ and taking expectations implies,

$$\left|f(\lambda) - \left(1 + i\lambda\mathbb{E}X - \frac{\lambda^2}{2!}\mathbb{E}\left[X^2\right]\right)\right| \le \mathbb{E}\left|e^{i\lambda X} - \left(1 + i\lambda X - \lambda^2\frac{X^2}{2!}\right)\right| \le \lambda^2 \mathbb{E}\left[X^2 \wedge \frac{|\lambda|\,|X|^3}{3!}\right] =: \lambda^2 \varepsilon(\lambda).$$

The DCT, with $X^2 \in L^1(P)$ being the dominating function, allows us to conclude that $\lim_{\varepsilon \to 0} \varepsilon(\lambda) = 0$. ∎

15

Weak Convergence of Random Sums

Throughout this chapter, we will assume the following standing notation unless otherwise stated. For each $n \in \mathbb{N}$, let $\{X_{n,k}\}_{k=1}^n$ be independent random variables and let

$$S_n := \sum_{k=1}^n X_{n,k}. \tag{15.1}$$

Until further notice we are going to assume $\mathbb{E}[X_{n,k}] = 0$, $\sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] < \infty$, and $\operatorname{Var}(S_n) = \sum_{k=1}^n \sigma_{n,k}^2 = 1$. Also let

$$f_{nk}(\lambda) := \mathbb{E}\left[e^{i\lambda X_{n,k}}\right] \tag{15.2}$$

denote the characteristic function of $X_{n,k}$.

Example 15.1. Suppose $\{X_n\}_{n=1}^\infty$ are mean zero square integrable random variables with $\sigma_k^2 = \operatorname{Var}(X_n)$. If we let $s_n^2 := \sum_{k=1}^n \operatorname{Var}(X_k) = \sum_{k=1}^n \sigma_k^2$, $\sigma_{n,k}^2 := \sigma_k^2 / s_n^2$, and $X_{n,k} := X_k / s_n$, then $\{X_{n,k}\}_{k=1}^n$ satisfy the above hypothesis and $S_n = \frac{1}{s_n} \sum_{k=1}^n X_k$.

Our main interest in this chapter is to consider the limiting behavior of S_n as $n \to \infty$. In order to do this, it will be useful to put conditions on the $\{X_{n,k}\}$ such that no one term dominates sum defining the sum defining S_n in Eq. (15.1) in the limit as $n \to \infty$.

Definition 15.2. *We say that $\{X_{n,k}\}$ satisfies the **Lindeberg Condition** (LC) iff*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > t\right] = 0 \text{ for all } t > 0. \tag{15.3}$$

We say $\{X_{n,k}\}$ satisfies condition (M) if

$$D_n := \max\left\{\sigma_{n,k}^2 : k \le n\right\} \to 0 \text{ as } n \to \infty, \tag{15.4}$$

*and we say $\{X_{n,k}\}$ is **uniformly asymptotic negligibility (UAN)** if for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} \max_{k\le n} P(|X_{n,k}| > \varepsilon) = 0. \tag{15.5}$$

Remark 15.3. The reader should observe that in order for condition (M) to hold in the setup in Example 15.1 it is necessary that $\lim_{n\to\infty} s_n^2 = \infty$.

Lemma 15.4. *Let us continue the notation in Example 15.1. Then* $\{X_{n,k} := X_k/s_n\}$ *satisfies* (LC) *if either of two conditions hold;*

1. $\{X_n\}_{n=1}^{\infty}$ *are i.i.d.*
2. *The* $\{X_n\}_{n=1}^{\infty}$ *satisfy* ***Liapunov condition****; there exists some* $\alpha > 2$ *such that*
$$\lim_{n\to\infty} \frac{\sum_{k=1}^n \mathbb{E}\left|X_k\right|^\alpha}{s_n^\alpha} = 0. \tag{15.6}$$

More generally, if $\{X_{n,k}\}$ *satisfies the Liapunov condition,*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \varphi\left(\left|X_{n,k}\right|\right)\right] = 0$$

where $\varphi : [0,\infty) \to [0,\infty)$ *is a non-decreasing function such that* $\varphi(t) > 0$ *for all* $t > 0$*, then* $\{X_{n,k}\}$ *satisfies* (LC).

Proof. 1. If $\{X_n\}_{n=1}^{\infty}$ are i.i.d., then $s_n = \sqrt{n}\sigma$ where $\sigma^2 = \mathbb{E}X_1^2$ and

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : \left|X_{n,k}\right| > t\right] &= \frac{1}{s_n^2}\sum_{k=1}^n \mathbb{E}\left[X_k^2 : \left|X_k\right| > s_n t\right] \\ &= \frac{1}{n\sigma^2}\sum_{k=1}^n \mathbb{E}\left[X_1^2 : \left|X_1\right| > \sqrt{n}\sigma t\right] \\ &= \frac{1}{\sigma^2}\mathbb{E}\left[X_1^2 : \left|X_1\right| > \sqrt{n}\sigma t\right]\end{aligned} \tag{15.7}$$

which, by DCT, tends to zero as $n \to \infty$.

2. Assuming Eq. (15.6), then for any $t > 0$,

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : \left|X_{n,k}\right| > t\right] &\le \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \left|\frac{X_{n,k}}{t}\right|^{\alpha-2} : \left|X_{n,k}\right| > t\right] \\ &\le \frac{1}{t^{\alpha-2}}\sum_{k=1}^n \mathbb{E}\left[\left|X_{n,k}\right|^\alpha\right] = \frac{1}{t^{\alpha-2}s_n^\alpha}\sum_{k=1}^n \mathbb{E}\left|X_k\right|^\alpha \to 0.\end{aligned}$$

For the last assertion, working as above we have

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : \left|X_{n,k}\right| > t\right] &\le \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \frac{\varphi\left(\left|X_{n,k}\right|\right)}{\varphi(t)} : \left|X_{n,k}\right| > t\right] \\ &\le \frac{1}{\varphi(t)}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 \varphi\left(\left|X_{n,k}\right|\right)\right] \to 0\end{aligned}$$

as $n \to \infty$. ∎

Lemma 15.5. *Let* $\{X_{n,k}\}_{n=1}^{\infty}$ *be as above, then* $(LC) \implies (M) \implies (UAN)$.

Proof. For $k \le n$,

$$\sigma_{n,k}^2 = \mathbb{E}\left[X_{n,k}^2\right] = \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|\le t}\right] + \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|>t}\right]$$
$$\le t^2 + \mathbb{E}\left[X_{n,k}^2 1_{|X_{n,k}|>t}\right] \le t^2 + \sum_{m=1}^n \mathbb{E}\left[X_{n,m}^2 1_{|X_{n,m}|>t}\right]$$

and therefore using (LC) we find

$$\lim_{n\to\infty} \max_{k\le n} \sigma_{n,k}^2 \le t^2 \text{ for all } t > 0.$$

This clearly implies (M) holds. The assertion that (M) implies (UAN) follows by Chebyschev's inequality,

$$\max_{k\le n} P\left(|X_{n,k}| > \varepsilon\right) \le \max_{k\le n} \frac{1}{\varepsilon^2}\mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| > \varepsilon\right]$$
$$\le \frac{1}{\varepsilon^2}\sum_{k=1}^n \mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| > \varepsilon\right] \to 0.$$

In fact the same argument shows that (M) implies

$$\sum_{k=1}^n P\left(|X_{n,k}| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\sum_{k=1}^n \mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| > \varepsilon\right] \to 0.$$

■

We will need the following lemma for our subsequent applications of the continuity theorem.

Lemma 15.6. *Suppose that $a_i, b_i \in \mathbb{C}$ with $|a_i|, |b_i| \le 1$ for $i = 1, 2, \dots, n$. Then*

$$\left|\prod_{i=1}^n a_i - \prod_{i=1}^n b_i\right| \le \sum_{i=1}^n |a_i - b_i|.$$

Proof. Let $a := \prod_{i=1}^{n-1} a_i$ and $b := \prod_{i=1}^{n-1} b_i$ and observe that $|a|, |b| \le 1$ and that

$$|a_n a - b_n b| \le |a_n a - a_n b| + |a_n b - b_n b|$$
$$= |a_n|\,|a - b| + |a_n - b_n|\,|b|$$
$$\le |a - b| + |a_n - b_n|.$$

The proof is now easily completed by induction on n. ■

Theorem 15.7 (Lindeberg-Feller CLT (I)). *Suppose $\{X_{n,k}\}$ satisfies (LC), then*

$$S_n \implies N(0,1). \tag{15.8}$$

(See Theorem 15.11 for a converse to this theorem.)

To prove this theorem we must show

$$\mathbb{E}\left[e^{i\lambda S_n}\right] \to e^{-\lambda^2/2} \text{ as } n \to \infty. \tag{15.9}$$

Before starting the formal proof, let me give an informal explanation for Eq. (15.9). Using

$$f_{nk}(\lambda) \sim 1 - \frac{\lambda^2}{2}\sigma_{nk}^2,$$

we might expect

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda S_n}\right] &= \prod_{k=1}^{n} f_{nk}(\lambda) = e^{\sum_{k=1}^{n} \ln f_{nk}(\lambda)} \\ &= e^{\sum_{k=1}^{n} \ln(1+f_{nk}(\lambda)-1)} \\ &\overset{(A)}{\sim} e^{\sum_{k=1}^{n}(f_{nk}(\lambda)-1)} \left(= \prod_{k=1}^{n} e^{(f_{nk}(\lambda)-1)}\right) \\ &\overset{(B)}{\sim} e^{\sum_{k=1}^{n} -\frac{\lambda^2}{2}\sigma_{nk}^2} = e^{-\frac{\lambda^2}{2}}.\end{aligned}$$

The question then becomes under what conditions are these approximations valid. It turns out that approximation (A), namely that

$$\lim_{n\to\infty}\left|\prod_{k=1}^{n} f_{nk}(\lambda) - \exp\left(\sum_{k=1}^{n}(f_{nk}(\lambda)-1)\right)\right| = 0, \tag{15.10}$$

is valid if condition (M) holds, see Lemma 15.9 below. It is shown in the estimate Eq. (15.11) below that the approximation (B) is valid, i.e.

$$\lim_{n\to\infty}\sum_{k=1}^{n}(f_{nk}(\lambda)-1) = -\frac{1}{2}\lambda^2,$$

if (LC) is satisfied. These observations would then constitute a proof of Theorem 15.7. The proof given below of Theorem 15.7 will not quite follow this route and will not use Lemma 15.9 directly. However, this lemma will be used in the proofs of Theorems 15.11 and 15.14.

Proof. Now on to the formal proof of Theorem 15.7. Since

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n} f_{nk}(\lambda) \text{ and } e^{-\lambda^2/2} = \prod_{k=1}^{n} e^{-\lambda^2\sigma_{n,k}^2/2},$$

we may use Lemma 15.6 to conclude,

$$\left|\mathbb{E}\left[e^{i\lambda S_n}\right] - e^{-\lambda^2/2}\right| \le \sum_{k=1}^{n}\left|f_{nk}(\lambda) - e^{-\lambda^2\sigma_{n,k}^2/2}\right| = \sum_{k=1}^{n}(A_{n,k} + B_{n,k})$$

where

$$A_{n,k} := \left| f_{nk}(\lambda) - \left[1 - \frac{\lambda^2 \sigma_{n,k}^2}{2}\right]\right| \text{ and}$$
$$B_{n,k} := \left| \left[1 - \frac{\lambda^2 \sigma_{n,k}^2}{2}\right] - e^{-\lambda^2 \sigma_{n,k}^2/2}\right|.$$

Now, using Lemma 14.47,

$$\begin{aligned}
A_{n,k} &= \left|\mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 + \frac{\lambda^2}{2} X_{n,k}^2\right]\right| \le \mathbb{E}\left|e^{i\lambda X_{n,k}} - 1 + \frac{\lambda^2}{2} X_{n,k}^2\right| \\
&\le \lambda^2 \mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda| |X_{n,k}|^3}{3!}\right] \\
&\le \lambda^2 \mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda| |X_{n,k}|^3}{3!} : |X_{n,k}| \le \varepsilon\right] + \lambda^2 \mathbb{E}\left[X_{n,k}^2 \wedge \frac{|\lambda| |X_{n,k}|^3}{3!} : |X_{n,k}| > \varepsilon\right] \\
&\le \lambda^2 \mathbb{E}\left[\frac{|\lambda| |X_{n,k}|^3}{3!} : |X_{n,k}| \le \varepsilon\right] + \lambda^2 \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right] \\
&\le \frac{\lambda^2}{3!} |\lambda| \varepsilon \cdot \mathbb{E}\left[|X_{n,k}|^2 : |X_{n,k}| \le \varepsilon\right] + \lambda^2 \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right] \\
&= \frac{|\lambda|^3 \varepsilon}{6} \sigma_{n,k}^2 + \lambda^2 \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right].
\end{aligned}$$

From this estimate and (LC) it follows that

$$\limsup_{n\to\infty} \sum_{k=1}^n A_{n,k} \le \limsup_{n\to\infty} \left(\frac{\lambda^3 \varepsilon}{6} + \lambda^2 \sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > \varepsilon\right]\right) = \frac{\lambda^3 \varepsilon}{6} \tag{15.11}$$

and since $\varepsilon > 0$ is arbitrary, we may conclude that $\limsup_{n\to\infty} \sum_{k=1}^n A_{n,k} = 0$.

To estimate $\sum_{k=1}^n B_{n,k}$, we use the estimate, $|e^{-u} - 1 + u| \le u^2/2$ valid for $u \ge 0$ (see Eq. 14.49 with $z = -u$). With this estimate we find,

$$\begin{aligned}
\sum_{k=1}^n B_{n,k} &= \sum_{k=1}^n \left| \left[1 - \frac{\lambda^2 \sigma_{n,k}^2}{2}\right] - e^{-\lambda^2 \sigma_{n,k}^2/2}\right| \\
&\le \sum_{k=1}^n \frac{1}{2} \left[\frac{\lambda^2 \sigma_{n,k}^2}{2}\right]^2 = \frac{\lambda^4}{8} \sum_{k=1}^n \sigma_{n,k}^4 \\
&\le \frac{\lambda^4}{8} \max_{k\le n} \sigma_{n,k}^2 \sum_{k=1}^n \sigma_{n,k}^2 = \frac{\lambda^4}{8} \max_{k\le n} \sigma_{n,k}^2 \to 0,
\end{aligned}$$

wherein we have used (M) (which is implied by (LC)) in taking the limit as $n \to \infty$. ■

As an application of Theorem 15.7 we can give half of the proof of Theorem 12.12.

Theorem 15.8 (Converse assertion in Theorem 12.12). *If $\{X_n\}_{n=1}^{\infty}$ are independent random variables and the random series, $\sum_{n=1}^{\infty} X_n$, is almost surely convergent, then for all $c > 0$ the following three series converge;*

1. $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$,
2. $\sum_{n=1}^{\infty} \mathrm{Var}\left(X_n 1_{|X_n| \le c}\right) < \infty$, *and*
3. $\sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n| \le c}\right)$ *converges.*

Proof. Since $\sum_{n=1}^{\infty} X_n$ is almost surely convergent, it follows that $\lim_{n\to\infty} X_n = 0$ a.s. and hence for every $c > 0$, $P(\{|X_n| \ge c \text{ i.o.}\}) = 0$. According the Borel zero one law this implies for every $c > 0$ that $\sum_{n=1}^{\infty} P(|X_n| > c) < \infty$. Since $X_n \to 0$ a.s., $\{X_n\}$ and $\left\{X_n^c := X_n 1_{|X_n| \le c}\right\}$ are tail equivalent for all $c > 0$. In particular $\sum_{n=1}^{\infty} X_n^c$ is almost surely convergent for all $c > 0$.

Fix $c > 0$, let $Y_n := X_n^c - \mathbb{E}[X_n^c]$ and let

$$s_n^2 = \mathrm{Var}(Y_1 + \cdots + Y_n) = \sum_{k=1}^{n} \mathrm{Var}(Y_k) = \sum_{k=1}^{n} \mathrm{Var}(X_k^c) = \sum_{k=1}^{n} \mathrm{Var}\left(X_k 1_{|X_k| \le c}\right).$$

For the sake of contradictions, suppose $s_n^2 \to \infty$ as $n \to \infty$. Since $|Y_k| \le 2c$, it follows that $\sum_{k=1}^{n} \mathbb{E}\left[Y_k^2 1_{|Y_k| > s_n t}\right] = 0$ for all sufficiently large n and hence

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[Y_k^2 1_{|Y_k| > s_n t}\right] = 0,$$

i.e. $\{Y_{n,k} := Y_k / s_n\}_{n=1}^{\infty}$ satisfies (LC) – see Examples 15.1 and Remark 15.3. So by the central limit Theorem 15.7, it follows that

$$\frac{1}{s_n^2} \sum_{k=1}^{n} (X_n^c - \mathbb{E}[X_n^c]) = \frac{1}{s_n^2} \sum_{k=1}^{n} Y_k \implies N(0,1).$$

On the other hand we know

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} X_n^c = \frac{\sum_{k=1}^{\infty} X_k^c}{\lim_{n\to\infty} s_n^2} = 0 \text{ a.s.}$$

and so by Slutsky's theorem,

$$\frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[X_n^c] = \frac{1}{s_n^2} \sum_{k=1}^{n} X_n^c - \frac{1}{s_n^2} \sum_{k=1}^{n} Y_k \implies N(0,1).$$

But it is not possible for constant (i.e. **non-random**) variables, $c_n := \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[X_n^c]$, to converge to a non-degenerate limit. (Think about this either in terms of characteristic functions or in terms of distribution functions.) Thus we must conclude that

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(X_n 1_{|X_n|\le c}\right) = \sum_{n=1}^{\infty} \operatorname{Var}\left(X_n^c\right) = \lim_{n\to\infty} s_n^2 < \infty.$$

An application of Kolmogorov's convergence criteria (Theorem 12.11) implies that

$$\sum_{n=1}^{\infty} \left(X_n^c - \mathbb{E}\left[X_n^c\right]\right) \text{ is convergent } a.s.$$

Since we already know that $\sum_{n=1}^{\infty} X_n^c$ is convergent almost surely we may now conclude $\sum_{n=1}^{\infty} \mathbb{E}\left(X_n 1_{|X_n|\le c}\right)$ is convergent. ■

Let us now turn to the converse of Theorem 15.7, see Theorem 15.11 below.

Lemma 15.9. *Suppose that* $\{X_{n,k}\}$ *satisfies property* (M), *i.e.* $D_n := \max_{k\le n} \sigma_{n,k}^2 \to 0$. *If we define,*

$$\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1 = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right],$$

then;

1. $\lim_{n\to\infty} \max_{k\le n} |\varphi_{n,k}(\lambda)| = 0$ *and*
2. $f_{S_n}(\lambda) - \prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)} \to 0$ *as* $n\to\infty$, *where*

$$f_{S_n}(\lambda) = \mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n} f_{n,k}(\lambda).$$

Proof. For the first item we estimate,

$$\begin{aligned}
\left|\mathbb{E}e^{i\lambda X} - 1\right| &\le \mathbb{E}\left|e^{i\lambda X} - 1\right| \le \mathbb{E}\left[2\wedge|\lambda X|\right] \\
&= \mathbb{E}\left[2\wedge|\lambda X| : |X|\ge\varepsilon\right] + \mathbb{E}\left[2\wedge|\lambda X| : |X|<\varepsilon\right] \\
&\le 2P\left[|X|\ge\varepsilon\right] + |\lambda|\varepsilon \le \frac{2}{\varepsilon^2}\mathbb{E}|X|^2 + |\lambda|\varepsilon
\end{aligned}$$

Replacing X by $X_{n,k}$ and in the above inequality shows

$$|\varphi_{n,k}(\lambda)| = |f_{n,k}(\lambda) - 1| \le \frac{2}{\varepsilon^2}\mathbb{E}|X_{n,k}|^2 + |\lambda|\varepsilon = \frac{2\sigma_{n,k}^2}{\varepsilon^2} + |\lambda|\varepsilon.$$

Therefore,

$$\limsup_{n\to\infty} \max_{k\le n} |\varphi_{n,k}(\lambda)| \le \limsup_{n\to\infty}\left[\frac{2D_n}{\varepsilon^2} + |\lambda|\varepsilon\right] = |\lambda|\varepsilon \to 0 \text{ as } \varepsilon\downarrow 0.$$

For the second item, observe that $\operatorname{Re}\varphi_{n,k}(\lambda) = \operatorname{Re} f_{n,k}(\lambda) - 1 \le 0$ and hence

$$\left|e^{\varphi_{n,k}(\lambda)}\right| = e^{\operatorname{Re}\varphi_{n,k}(\lambda)} \le e^0 = 1$$

and hence we have from Lemma 15.6 and the estimate (14.49),

$$\begin{aligned}\left|\prod_{k=1}^{n} f_{n,k}(\lambda)-\prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)}\right| &\leq \sum_{k=1}^{n}\left|f_{n,k}(\lambda)-e^{\varphi_{n,k}(\lambda)}\right| \\ &=\sum_{k=1}^{n}\left|e^{\varphi_{n,k}(\lambda)}-1-\varphi_{n,k}(\lambda)\right| \\ &\leq \frac{1}{2}\sum_{k=1}^{n}\left|\varphi_{n,k}(\lambda)\right|^{2} \\ &\leq \frac{1}{2}\max_{k\leq n}\left|\varphi_{n,k}(\lambda)\right|\cdot\sum_{k=1}^{n}\left|\varphi_{n,k}(\lambda)\right|.\end{aligned}$$

Moreover since $\mathbb{E}X_{n,k}=0$, the estimate in Eq. (14.49) implies

$$\begin{aligned}\sum_{k=1}^{n}\left|\varphi_{n,k}(\lambda)\right| &=\sum_{k=1}^{n}\left|\mathbb{E}\left[e^{i\lambda X_{n,k}}-1-i\lambda X_{n,k}\right]\right| \\ &\leq \sum_{k=1}^{n}\left|\mathbb{E}\left[\frac{1}{2}\left|\lambda X_{n,k}\right|^{2}\right]\right| \leq \frac{\lambda^{2}}{2}\sum_{k=1}^{n}\sigma_{n,k}^{2}=\frac{\lambda^{2}}{2}.\end{aligned}$$

Thus we have shown,

$$\left|\prod_{k=1}^{n} f_{n,k}(\lambda)-\prod_{k=1}^{n} e^{\varphi_{n,k}(\lambda)}\right| \leq \frac{\lambda^{2}}{4}\max_{k\leq n}\left|\varphi_{n,k}(\lambda)\right|$$

and the latter expression tends to zero by item 1. ■

Lemma 15.10. *Let X be a random variable such that $\mathbb{E}X^2<\infty$ and $\mathbb{E}X=0$. Further let $f(\lambda):=\mathbb{E}\left[e^{i\lambda X}\right]$ and $u(\lambda):=\operatorname{Re}(f(\lambda)-1)$. Then for all $c>0$,*

$$u(\lambda)+\frac{\lambda^{2}}{2}\mathbb{E}\left[X^{2}\right]\geq \mathbb{E}\left[X^{2}\left[\frac{\lambda^{2}}{2}-\frac{2}{c^{2}}\right]:|X|>c\right] \tag{15.12}$$

or equivalently

$$\mathbb{E}\left[\cos\lambda X-1+\frac{\lambda^{2}}{2}X^{2}\right]\geq \mathbb{E}\left[X^{2}\left[\frac{\lambda^{2}}{2}-\frac{2}{c^{2}}\right]:|X|>c\right]. \tag{15.13}$$

In particular if we choose $|\lambda|\geq\sqrt{6}/|c|$, then

$$\mathbb{E}\left[\cos\lambda X-1+\frac{\lambda^{2}}{2}X^{2}\right]\geq \frac{1}{c^{2}}\mathbb{E}\left[X^{2}:|X|>c\right]. \tag{15.14}$$

Proof. For all $\lambda\in\mathbb{R}$, we have (see Eq. (14.48)) $\cos\lambda X-1+\frac{\lambda^2}{2}X^2\geq 0$ and $\cos\lambda X-1\geq -2$. Therefore,

$$\begin{aligned} u(\lambda) + \frac{\lambda^2}{2}\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2\right] \\ &\geq \mathbb{E}\left[\cos\lambda X - 1 + \frac{\lambda^2}{2}X^2 : |X| > c\right] \\ &\geq \mathbb{E}\left[-2 + \frac{\lambda^2}{2}X^2 : |X| > c\right] \\ &\geq \mathbb{E}\left[-2\frac{|X|^2}{c^2} + \frac{\lambda^2}{2}X^2 : |X| > c\right] \end{aligned}$$

which gives Eq. (15.12). ■

Theorem 15.11 (Lindeberg-Feller CLT (II)). *Suppose $\{X_{n,k}\}$ satisfies (M) and also the central limit theorem in Eq. (15.8) holds, then $\{X_{n,k}\}$ satisfies (LC). So under condition (M), S_n converges to a normal random variable iff (LC) holds.*

Proof. By assumption we have

$$\lim_{n\to\infty}\max_{k\leq n}\sigma_{n,k}^2 = 0 \text{ and } \lim_{n\to\infty}\prod_{k=1}^n f_{n,k}(\lambda) = e^{-\lambda^2/2}.$$

The second inequality combined with Lemma 15.9 implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^n \varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\prod_{k=1}^n e^{\varphi_{n,k}(\lambda)} = e^{-\lambda^2/2}.$$

Taking the modulus of this equation then implies,

$$\lim_{n\to\infty} e^{\sum_{k=1}^n \operatorname{Re}\varphi_{n,k}(\lambda)} = \lim_{n\to\infty}\left|e^{\sum_{k=1}^n \varphi_{n,k}(\lambda)}\right| = e^{-\lambda^2/2}$$

from which we may conclude

$$\lim_{n\to\infty}\sum_{k=1}^n \operatorname{Re}\varphi_{n,k}(\lambda) = -\lambda^2/2.$$

We may write this last limit as

$$\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\left[\cos(\lambda X_{n,k}) - 1 + \frac{\lambda^2}{2}X_{n,k}^2\right] = 0$$

which by Lemma 15.10 implies

$$\lim_{n\to\infty}\sum_{k=1}^n \mathbb{E}\left[X_{n,k}^2 : |X_{n,k}| > c\right] = 0$$

for all $c > 0$ which is (LC). ■

15.1 Infinitely Divisible and Stable Symmetric Distributions

To get some indication as to what we might expect to happen when the Lindeberg condition is relaxed, we consider the following Poisson limit theorem.

Theorem 15.12 (A Poisson Limit Theorem). *For each* $n \in \mathbb{N}$, *let* $\{X_{n,k}\}_{k=1}^n$ *be independent Bernoulli random variables with* $P(X_{n,k}=1) = p_{n,k}$ *and* $P(X_{n,k}=0) = q_{n,k} := 1 - p_{n,k}$. *Suppose;*

1. $\lim_{n\to\infty} \sum_{k=1}^n p_{n,k} = a \in (0,\infty)$ *and*
2. $\lim_{n\to\infty} \max_{1\le k\le n} p_{n,k} = 0$. *(So no one term is dominating the sums in item 1.)*

Then $S_n = \sum_{k=1}^n X_{n,k} \implies Z$ *where* Z *is a Poisson random variable with mean* a. *(See Section 2.6 of Durrett[*__?__*] for more on this theorem.)*

Proof. Recall from Example 14.14 that for any $a > 0$,

$$\mathbb{E}\left[e^{i\lambda Z}\right] = \exp\left(a\left(e^{i\lambda}-1\right)\right).$$

Since

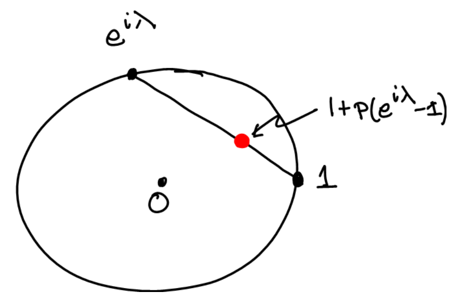
$$\mathbb{E}\left[e^{i\lambda X_{n,k}}\right] = e^{i\lambda}p_{n,k} + (1-p_{n,k}) = 1 + p_{n,k}\left(e^{i\lambda}-1\right),$$

it follows that

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^n \left[1 + p_{n,k}\left(e^{i\lambda}-1\right)\right].$$

Since $1+p_{n,k}\left(e^{i\lambda}-1\right)$ lies on the line segment joining 1 to $e^{i\lambda}$, it follows that

$$\left|1 + p_{n,k}\left(e^{i\lambda}-1\right)\right| \le 1.$$



Hence we may apply Lemma 15.6 to find

$$\begin{aligned}&\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&\leq\sum_{k=1}^{n}\left|\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&=\sum_{k=1}^{n}\left|\exp\left(z_{n,k}\right)-\left[1+z_{n,k}\right]\right|\end{aligned}$$

where

$$z_{n,k}=p_{n,k}\left(e^{i\lambda}-1\right).$$

Since $\operatorname{Re} z_{n,k}=p_{n,k}\left(\cos\lambda-1\right)\leq 0$, we may use the calculus estimate in Eq. (14.49) to conclude,

$$\begin{aligned}&\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\\&\leq\frac{1}{2}\sum_{k=1}^{n}\left|z_{n,k}\right|^{2}\leq\frac{1}{2}\max_{1\leq k\leq n}\left|z_{n,k}\right|\sum_{k=1}^{n}\left|z_{n,k}\right|\\&\leq 2\max_{1\leq k\leq n}p_{n,k}\sum_{k=1}^{n}p_{n,k}.\end{aligned}$$

Using the assumptions, we may conclude

$$\left|\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)-\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\right|\to 0\text{ as }n\to\infty.$$

Since

$$\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)=\exp\left(\sum_{k=1}^{n}p_{n,k}\left(e^{i\lambda}-1\right)\right)\to\exp\left(a\left(e^{i\lambda}-1\right)\right),$$

we have shown

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[e^{i\lambda S_n}\right]&=\lim_{n\to\infty}\prod_{k=1}^{n}\left[1+p_{n,k}\left(e^{i\lambda}-1\right)\right]\\&=\lim_{n\to\infty}\prod_{k=1}^{n}\exp\left(p_{n,k}\left(e^{i\lambda}-1\right)\right)=\exp\left(a\left(e^{i\lambda}-1\right)\right).\end{aligned}$$

The result now follows by an application of the continuity Theorem 14.21. ■

Remark 15.13. Keeping the notation in Theorem 15.12, we have

$$\mathbb{E}\left[X_{n,k}\right]=p_{n,k}\text{ and }\operatorname{Var}\left(X_{n,k}\right)=p_{n,k}\left(1-p_{n,k}\right)$$

and

$$s_n^2 := \sum_{k=1}^{n} \operatorname{Var}(X_{n,k}) = \sum_{k=1}^{n} p_{n,k}(1 - p_{n,k}).$$

Under the assumptions of Theorem 15.12, we see that $s_n^2 \to a$ as $n \to \infty$. Let $Y_{n,k} := \frac{X_{n,k} - p_{n,k}}{s_n}$ so that $\mathbb{E}[Y_{n,k}] = 0$ and $\sigma_{n,k}^2 := \operatorname{Var}(Y_{n,k}) = \frac{1}{s_n^2} \operatorname{Var}(X_{n,k}) = \frac{1}{s_n^2} p_{n,k}(1 - p_{n,k})$ which satisfies condition (M). Let us observe that, for large n,

$$\begin{aligned}
\mathbb{E}\left[Y_{n,k}^2 : |Y_{n,k}| > t\right] &= \mathbb{E}\left[Y_{n,k}^2 : \left|\frac{X_{n,k} - p_{n,k}}{s_n}\right| > t\right] \\
&= \mathbb{E}\left[Y_{n,k}^2 : |X_{n,k} - p_{n,k}| > s_n t\right] \\
&\geq \mathbb{E}\left[Y_{n,k}^2 : |X_{n,k} - p_{n,k}| > 2at\right] \\
&= \mathbb{E}\left[Y_{n,k}^2 : X_{n,k} = 1\right] = p_{n,k}\left(\frac{1 - p_{n,k}}{s_n}\right)^2
\end{aligned}$$

from which it follows that

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[Y_{n,k}^2 : |Y_{n,k}| > t\right] = \lim_{n\to\infty} \sum_{k=1}^{n} p_{n,k}\left(\frac{1 - p_{n,k}}{s_n}\right)^2 = a.$$

Therefore $\{Y_{n,k}\}$ do not satisfy (LC). Nevertheless we have

$$\sum_{k=1}^{n} Y_{n,k} = \frac{\sum_{k=1}^{n} X_{n,k} - \sum_{k=1}^{n} p_{n,k}}{s_n} \implies \frac{Z - a}{a}$$

where Z is a Poisson random variable with mean a. Notice that the limit is **not** a normal random variable.

We wish to characterize the possible limiting distributions of sequences $\{S_n\}_{n=1}^{\infty}$ when we relax the Lindeberg condition (LC) to condition (M). We have the following theorem.

Theorem 15.14. *Suppose* $\{X_{n,k}\}_{k=1}^{n}$ *satisfy property* (M) *and* $S_n := \sum_{k=1}^{n} X_{n,k} \implies L$ *for some random variable* L. *Then the characteristic function* $f_L(\lambda) := \mathbb{E}\left[e^{i\lambda L}\right]$ *must be of the form,*

$$f_L(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

where ν *– is a finite positive measure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *such that* $\nu(\mathbb{R}) \leq 1$. *(Recall that you proved in Exercise 14.4 that* $\exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$ *is always the characteristic function of a probability measure.)*

Proof. As before, let $f_{n,k}(\lambda) = \mathbb{E}\left[e^{i\lambda X_{n,k}}\right]$ and $\varphi_{n,k}(\lambda) := f_{n,k}(\lambda) - 1$. By the continuity theorem we are assuming

$$\lim_{n\to\infty} f_{S_n}(\lambda) = \lim_{n\to\infty} \prod_{k=1}^{n} f_{n,k}(\lambda) = f(\lambda)$$

where $f(\lambda)$ is continuous at $\lambda = 0$. We are also assuming property (M), i.e.

$$\lim_{n\to\infty} \max_{k\le n} \sigma_{n,k}^2 = 0.$$

Under condition (M), we expect $f_{n,k}(\lambda) \cong 1$ for n large. Therefore we expect

$$f_{n,k}(\lambda) = e^{\ln f_{n,k}(\lambda)} = e^{\ln[1+(f_{n,k}(\lambda)-1)]} \cong e^{(f_{n,k}(\lambda)-1)}$$

and hence that

$$\mathbb{E}\left[e^{i\lambda S_n}\right] = \prod_{k=1}^{n} f_{n,k}(\lambda) \cong \prod_{k=1}^{n} e^{(f_{n,k}(\lambda)-1)} = \exp\left(\sum_{k=1}^{n}(f_{n,k}(\lambda)-1)\right). \tag{15.15}$$

This is in fact correct, since Lemma 15.9 indeed implies

$$\lim_{n\to\infty}\left[\mathbb{E}\left[e^{i\lambda S_n}\right] - \exp\left(\sum_{k=1}^{n}(f_{n,k}(\lambda)-1)\right)\right] = 0. \tag{15.16}$$

Since $\mathbb{E}[X_{n,k}] = 0$,

$$\begin{aligned} f_{n,k}(\lambda) - 1 &= \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1\right] = \mathbb{E}\left[e^{i\lambda X_{n,k}} - 1 - i\lambda X_{n,k}\right] \\ &= \int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x) \end{aligned}$$

where $\mu_{n,k} := P \circ X_{n,k}^{-1}$ is the law of $X_{n,k}$. Therefore we have

$$\begin{aligned} \exp\left(\sum_{k=1}^{n}(f_{n,k}(\lambda)-1)\right) &= \exp\left(\sum_{k=1}^{n}\int_{\mathbb{R}}\left(e^{i\lambda x} - 1 - i\lambda x\right) d\mu_{n,k}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}}\left(e^{i\lambda x} - 1 - i\lambda x\right)\sum_{k=1}^{n} d\mu_{n,k}(x)\right) \\ &= \exp\left(\int_{\mathbb{R}}\left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_n^*(x)\right) \end{aligned} \tag{15.17}$$

where $\nu_n^* := \sum_{k=1}^{n} \mu_{n,k}$. Let us further observe that

$$\int_{\mathbb{R}} x^2 d\nu_n^*(x) = \sum_{k=1}^{n}\int_{\mathbb{R}} x^2 d\mu_{n,k}(x) = \sum_{k=1}^{n}\sigma_{n,k}^2 = 1.$$

Hence if we define $d\nu^*(x) := x^2 d\nu_n^*(x)$, then ν_n is a probability measure and we have from Eqs. (15.16) and Eq. (15.17) that

$$\left| f_{S_n}(\lambda) - \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_n(x)\right)\right| \to 0. \tag{15.18}$$

Since $h(x) := \frac{e^{i\lambda x}-1-i\lambda x}{x^2}$ is a continuous function of $\bar{\mathbb{R}}$ with $h(\pm\infty) = 0$, there is a subsequence, $\{n_l\}$ of $\{n\}$ such that $\nu_{n_l}(h) \to \bar{\nu}(h)$ for some probability measure on $\left(\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}\right)$. Combining this with Eq. (15.18) allows us to conclude,

$$\begin{aligned} f_L(\lambda) &= \lim_{l\to\infty} \mathbb{E}\left[e^{i\lambda S_{n_l}}\right] = \lim_{l\to\infty} \exp\left(\int_{\mathbb{R}} \left(e^{i\lambda x} - 1 - i\lambda x\right) d\nu_{n_l}^*(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right). \end{aligned}$$

■

Definition 15.15. *We say that $\{X_{n,k}\}_{k=1}^n$ has bounded variation (BV) iff*

$$\sup_n \operatorname{Var}(S_n) = \sup_n \sum_{k=1}^n \sigma_{n,k}^2 < \infty. \tag{15.19}$$

Corollary 15.16. *Suppose $\{X_{n,k}\}_{k=1}^n$ satisfy properties (M) and (BV). If $S_n := \sum_{k=1}^n X_{n,k} \implies L$ for some random variable L, then*

$$f_L(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right) \tag{15.20}$$

where ν – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Proof. Let $s_n^2 := \operatorname{Var}(S_n)$. If $\lim_{n\to\infty} s_n = 0$, then $S_n \to 0$ in L^2 and hence weakly, therefore Eq. (15.20) holds with $\nu \equiv 0$. So let us now suppose $\lim_{n\to\infty} s_n \neq 0$. Since $\{s_n\}_{n=1}^\infty$ is bounded, we may by passing to a subsequence if necessary, assume $\lim_{n\to\infty} s_n = s > 0$. By replacing $X_{n,k}$ by $X_{n,k}/s_n$ and hence S_n by S_n/s_n, we then know by Slutsky's theorem that $S_n/s_n \implies L/s$. Hence by an application of Theorem 15.14, we may conclude

$$f_L(\lambda/s) = f_{L/s}(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

where ν – is a finite positive measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\nu(\mathbb{R}) \le 1$. Letting $\lambda \to s\lambda$ in this expression then implies

$$\begin{aligned} f_L(\lambda) &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{x^2} d\nu(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda sx} - 1 - i\lambda sx}{(sx)^2} s^2 d\nu(x)\right) \\ &= \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu_s(x)\right) \end{aligned}$$

where ν_s is the finite measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ defined by

$$\nu_s(A) := s^2\nu\left(s^{-1}A\right) \text{ for all } A \in \mathcal{B}_{\mathbb{R}}.$$

■

The reader should observe that

$$\frac{e^{i\lambda x}-1-i\lambda x}{x^2} = \frac{1}{x^2}\sum_{k=2}^{\infty}\frac{(i\lambda x)^k}{k!} = \frac{1}{x^2}\sum_{k=2}^{\infty}\frac{i^k}{k!}\lambda^k x^{k-2}$$

and hence $(\lambda, x) \to \frac{e^{i\lambda x}-1-i\lambda x}{x^2}$ is smooth. Moreover,

$$\frac{d}{d\lambda}\frac{e^{i\lambda x}-1-i\lambda x}{x^2} = \frac{ixe^{i\lambda x}-ix}{x^2} = i\frac{e^{i\lambda x}-1}{x}$$

and

$$\frac{d^2}{d\lambda^2}\frac{e^{i\lambda x}-1-i\lambda x}{x^2} = i\frac{ixe^{i\lambda x}}{x} = -e^{i\lambda x}.$$

Using these remarks and the fact that $\nu(\mathbb{R}) < \infty$, it is easy to see that

$$f_L'(\lambda) = \left(\int_{\mathbb{R}} i\frac{e^{i\lambda x}-1}{x}d\nu_s(x)\right)f_L(\lambda)$$

and

$$f_L''(\lambda) = \left(\int_{\mathbb{R}} -e^{i\lambda x}d\nu_s(x) + \left[\left(\int_{\mathbb{R}} i\frac{e^{i\lambda x}-1}{x}d\nu_s(x)\right)^2\right]\right)f_L(\lambda)$$

and in particular, $f_L'(0) = 0$ and $f_L''(0) = -\nu_s(\mathbb{R})$. Therefore the probability measure, μ, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\hat{\mu}(\lambda) = f_L(\lambda)$ has mean zero and variance, $\nu_s(\mathbb{R}) < \infty$.

Definition 15.17. *A probability distribution, μ, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ is* ***infinitely divisible*** *iff for all $n \in \mathbb{N}$ there exists i.i.d. nondegenerate random variables, $\{X_{n,k}\}_{k=1}^n$, such that $X_{n,1} + \cdots + X_{n,n} \overset{d}{=} \mu$. This can be formulated in the following two equivalent ways. For all $n \in \mathbb{N}$ there should exists a non-degenerate probability measure, μ_n, on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ such that $\mu_n^{*n} = \mu$. For all $n \in \mathbb{N}$, $\hat{\mu}(\lambda) = [g(\lambda)]^n$ for some non-constant characteristic function, g.*

Theorem 15.18. *The following class of symmetric distributions on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ are equal;*

1. *C_1 – all possible limiting distributions under properties (M) and (BV).*
2. *C_2 – all distributions with characteristic functions of the form given in Corollary 15.16.*
3. *C_3 – all infinitely divisible distributions with mean zero and finite variance.*

Proof. The inclusion, $C_1 \subset C_2$, is the content of Corollary 15.16. For $C_2 \subset C_3$, observe that if

$$\hat{\mu}(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} d\nu(x)\right)$$

then $\hat{\mu}(\lambda) = [\hat{\mu}_n(\lambda)]^n$ where μ_n is the unique probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that

$$\hat{\mu}_n(\lambda) = \exp\left(\int_{\mathbb{R}} \frac{e^{i\lambda x} - 1 - i\lambda x}{x^2} \frac{1}{n} d\nu(x)\right).$$

For $C_3 \subset C_1$, simply define $\{X_{n,k}\}_{k=1}^n$ to be i.i.d with $\mathbb{E}\left[e^{i\lambda X_{n,k}}\right] = \hat{\mu}_n(\lambda)$. In this case $S_n = \sum_{k=1}^n X_{n,k} \stackrel{d}{=} \mu$. ■

15.1.1 Stable Laws

See the file, dynkin-stable-infinitely-divs.pdf, and Durrett [**?**, Example 3.10 on p. 106 and Section 2.7.].

Part IV

Conditional Expectations and Martingales

16

Hilbert Space Basics

Definition 16.1. *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,

$$\begin{aligned}\|x + y\|^2 &= \langle x + y|x + y\rangle = \|x\|^2 + \|y\|^2 + \langle x|y\rangle + \langle y|x\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle.\end{aligned} \tag{16.1}$$

Theorem 16.2 (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 z is the "orthogonal projection" of x onto y, see Figure 16.1.) 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 $x = \|y\|^{-2}\langle x|y\rangle y$. ■

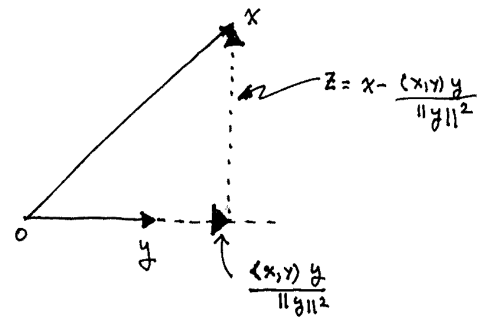


Fig. 16.1. The picture behind the proof of the Schwarz inequality.

Corollary 16.3. *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,

$$\begin{aligned}\|x+y\|^2 &= \|x\|^2 + \|y\|^2 + 2\mathrm{Re}\langle x|y\rangle \\ &\le \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2.\end{aligned}$$

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 $x, x', y, y' \in H$, then

$$\begin{aligned}|\langle x + \Delta x|y + \Delta y\rangle - \langle x|y\rangle| &= |\langle x|\Delta y\rangle + \langle \Delta x|y\rangle + \langle \Delta x|\Delta y\rangle| \\ &\le \|x\|\|\Delta y\| + \|y\|\|\Delta x\| + \|\Delta x\|\|\Delta y\| \\ &\to 0 \text{ as } \Delta x, \Delta y \to 0,\end{aligned}$$

from which it follows that $\langle\cdot|\cdot\rangle$ is continuous. ∎

Definition 16.4. *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 16.5. *Let $(H, \langle\cdot|\cdot\rangle)$ be an inner product space then*

1. *(**Parallelogram Law**)*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{16.2}$$

for all $x, y \in H$.

*2. (**Pythagorean Theorem**) If $S \subset\subset H$ is a finite orthogonal set, then*

$$\left\|\sum_{x\in S} x\right\|^2 = \sum_{x\in S} \|x\|^2. \tag{16.3}$$

*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 16.6. *A **Hilbert space** is an inner product space $(H, \langle\cdot|\cdot\rangle)$ such that the induced Hilbertian norm is complete.*

Example 16.7. 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 11.17 for the completeness assertion.

Definition 16.8. *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 16.9 (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 $(x-y) \perp M$.

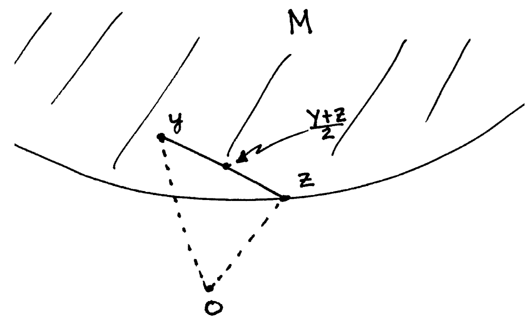


Fig. 16.2. The geometry of convex sets.

Proof. Uniqueness. By replacing M by $M - x := \{m - x : m \in M\}$ we may assume $x = 0$. Let $\delta := d(0, M) = \inf_{m \in M} \|m\|$ and $y, z \in M$, see Figure 16.2.

By the parallelogram law and the convexity of M,

$$2\|y\|^2 + 2\|z\|^2 = \|y + z\|^2 + \|y - z\|^2$$
$$= 4\left\|\frac{y + z}{2}\right\|^2 + \|y - z\|^2 \geq 4\delta^2 + \|y - z\|^2. \tag{16.4}$$

Hence if $\|y\| = \|z\| = \delta$, then $2\delta^2 + 2\delta^2 \geq 4\delta^2 + \|y - z\|^2$, so that $\|y - z\|^2 = 0$. Therefore, if a minimizer for $d(0, \cdot)|_M$ exists, it is unique.

Existence. Let $y_n \in M$ be chosen such that $\|y_n\| = \delta_n \to \delta \equiv d(0, M)$. Taking $y = y_m$ and $z = y_n$ in Eq. (16.4) shows

$$2\delta_m^2 + 2\delta_n^2 \geq 4\delta^2 + \|y_n - y_m\|^2.$$

Passing to the limit $m, n \to \infty$ in this equation implies,

$$2\delta^2 + 2\delta^2 \geq 4\delta^2 + \limsup_{m,n\to\infty} \|y_n - y_m\|^2,$$

i.e. $\limsup_{m,n\to\infty} \|y_n - y_m\|^2 = 0$. 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,

$$\|y\| = \lim_{n\to\infty} \|y_n\| = \delta = d(0, M).$$

So y is the desired point in M which is closest to 0.

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 $(x - y) \perp M$.

Finally suppose $y \in M$ is any point such that $(x-y) \perp M$. Then for $z \in M$, by Pythagorean's theorem,

$$\|x-z\|^2 = \|x-y+y-z\|^2 = \|x-y\|^2 + \|y-z\|^2 \geq \|x-y\|^2$$

which shows $d(x,M)^2 \geq \|x-y\|^2$. That is to say y is the point in M closest to x. ■

Definition 16.10. *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 16.15 below.) A bounded operator $A : H \to H$ is* ***self - adjoint*** *or* ***Hermitian*** *if $A = A^*$.*

Definition 16.11. *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*

$$\langle x|m\rangle = \langle P_M(x)|m\rangle \text{ for all } m \in M. \tag{16.5}$$

Theorem 16.12 (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 16.13. *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 16.1. Suppose M is a subset of H, then $M^{\perp\perp} = \overline{\mathrm{span}(M)}$.

Theorem 16.14 (Riesz Theorem). *Let H^* be the dual space of H (i.e. that linear space of continuous linear functionals on H). The map*

$$z \in H \xrightarrow{j} \langle \cdot|z\rangle \in H^* \tag{16.6}$$

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, with out loss of generality, is non-zero. Then $M =\mathrm{Nul}(f)$ – a closed proper subspace of H. Since, by Corollary 16.13, $H = M \oplus M^{\perp}$, $f : H/M \cong M^{\perp} \to \mathbb{F}$ 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 16.15 (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*

$$\langle Ax|y\rangle_K = \langle x|A^*y\rangle_H \text{ for all } x \in H \text{ and } y \in K. \tag{16.7}$$

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 $(A^*)^{-1} = \left(A^{-1}\right)^*$.*

Proof. For each $y \in K$, the map $x \to \langle Ax|y\rangle_K$ is in H^* and therefore there exists, by Theorem 16.14, 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 16.2.

Items 3. and 4. Making use of Schwarz's inequality (Theorem 16.2), 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

$$\|A^*A\| \leq \|A^*\| \|A\| = \|A\|^2$$

and

$$\begin{aligned}\|A\|^2 &= \sup_{h\in H:\|h\|=1} \|Ah\|^2 = \sup_{h\in H:\|h\|=1} |\langle Ah|Ah\rangle| \\ &= \sup_{h\in H:\|h\|=1} |\langle h|A^*Ah\rangle| \leq \sup_{h\in H:\|h\|=1} \|A^*Ah\| = \|A^*A\| \end{aligned} \tag{16.8}$$

we also have $\|A^*A\| \leq \|A\|^2 \leq \|A^*A\|$ which shows $\|A\|^2 = \|A^*A\|$.

Alternatively, from Eq. (16.8),

$$\|A\|^2 \leq \|A^*A\| \leq \|A\| \|A^*\| \tag{16.9}$$

which then implies $\|A\| \leq \|A^*\|$. Replacing A by A^* in this last inequality shows $\|A^*\| \leq \|A\|$ and hence that $\|A^*\| = \|A\|$. Using this identity back in Eq. (16.9) 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 16.2. 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 16.3. 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 16.16. *Suppose $A : H \to K$ is a bounded operator, then:*

1. $\text{Nul}(A^*) = \text{Ran}(A)^\perp$.
2. $\overline{\text{Ran}(A)} = \text{Nul}(A^*)^\perp$.
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 16.1, $\overline{\mathrm{Ran}(A)} = \mathrm{Ran}(A)^{\perp\perp}$, 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 16.17 (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. ■

16.1 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 compactess result.

Theorem 16.18. *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 $y_k \overset{w}{\to} x$.*

[3] A Banach space is a complete normed space. The main examples for us are Hilbert spaces.

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 16.14, 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

$$\lim_{k\to\infty} \langle y_k|z\rangle = \langle x|z\rangle \text{ for all } z \in H_0. \tag{16.10}$$

To finish the proof we need only observe that Eq. (16.10) 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, the previous states that a set is sequentially precompact in H iff it is bounded. Let us now use Theorem 16.18 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 16.19. *If $\{X_n\}_{n=1}^\infty \subset L^1(\Omega, \mathcal{B}, P)$ 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*

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in \mathcal{B}_b. \tag{16.11}$$

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 Theorem16.18, $\{X_n^m\}_{n=1}^\infty$ 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 $Y_k^m \xrightarrow{w} X^m$ 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. (16.11) 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 16.20. 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 $Y_k \xrightarrow{w} X$. Then for $h \in \mathcal{B}_b$ and any $\varepsilon > 0$ we would have

$$\mathbb{E}\left[X h 1_{(\varepsilon,1)}\right] = \lim_{k\to\infty} \mathbb{E}\left[Y_k h 1_{(\varepsilon,1)}\right] = 0.$$

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 16.19.

When $1 < p < \infty$, the situation is simpler.

Theorem 16.21. *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*

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[Xh] \text{ for all } h \in L^q(\Omega, \mathcal{B}, P). \tag{16.12}$$

Proof. Let $C := \sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$ and recall that Lemma 11.35 guarantees that $\{X_n\}_{n=1}^\infty$ is a uniformly integrable subset of $L^1(\Omega, \mathcal{B}, P)$. Therefore by Theorem 16.19, there exists $X \in L^1(\Omega, \mathcal{B}, P)$ and a subsequence, $Y_k := X_{n_k}$, such that Eq. (16.11) holds. We will complete the proof by showing; a) $X \in L^p(\Omega, \mathcal{B}, P)$ and b) and Eq. (16.12) 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 $h = \overline{\mathrm{sgn}(X)}\,|X|^{p-1} 1_{|X|\le M}$ in the previous inequality shows

$$\begin{aligned}\mathbb{E}\left[|X|^p 1_{|X|\le M}\right] &\le C \left\|\overline{\mathrm{sgn}(X)}\,|X|^{p-1} 1_{|X|\le M}\right\|_q \\ &= C\left(\mathbb{E}\left[|X|^{(p-1)q} 1_{|X|\le M}\right]\right)^{1/q} \le C\left(\mathbb{E}\left[|X|^p 1_{|X|\le M}\right]\right)^{1/q}\end{aligned}$$

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,

$$\limsup_{k\to\infty} |\mathbb{E}[(X - Y_k) h]| \le 2C \left\| h 1_{|h|>M} \right\|_q .$$

This completes the proof, since $\left\| h 1_{|h|>M} \right\|_q \to 0$ as $M \to \infty$ by DCT. ■

16.2 Exercises

Exercise 16.4. 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 (16.4). 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 16.5 (The Mean Ergodic Theorem). Let $U : H \to H$ be a unitary operator 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 $x \in H$.

Hints: 1. Show H is the orthogonal direct sum of M and $\overline{\mathrm{Ran}(U - I)}$ by first showing $\mathrm{Nul}(U^* - I) = \mathrm{Nul}(U - I)$ and then using Lemma 16.16. 2. Verify the result for $x \in \mathrm{Nul}(U - I)$ and $x \in \mathrm{Ran}(U - I)$. 3. Use a limiting argument to verify the result for $x \in \overline{\mathrm{Ran}(U - I)}$.

Solution to Exercise (16.5). Let $M = \mathrm{Nul}(U - I)$, then $S_n x = x$ for all $x \in M$. Notice that $x \in \mathrm{Nul}(U^* - I)$ iff $x = U^* x$ iff $Ux = UU^* x = x$, iff $x \in \mathrm{Nul}(U - I) = M$. Therefore

$$\overline{\operatorname{Ran}(U-I)} = \operatorname{Nul}(U^*-I)^\perp = \operatorname{Nul}(U-I)^\perp = M^\perp.$$

Suppose that $x = Uy - y \in \operatorname{Ran}(U-I)$ for some $y \in H$, then

$$S_n x = \frac{1}{n}\left(U^n y - y\right) \to 0 \text{ as } n \to \infty.$$

Finally if $x \in M^\perp$ and $y \in \operatorname{Ran}(U-I)$, we have

$$\|S_n x - S_n y\| \le \|x - y\|$$

and hence

$$\limsup_{n\to\infty} \|S_n x - S_n y\| \le \|x - y\|$$

from which 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

$$\lim_{n\to\infty} S_n x = \lim_{n\to\infty} S_n m + \lim_{n\to\infty} S_n m^\perp = m + 0 = P_M x.$$

17

The Radon-Nikodym Theorem

Theorem 17.1 (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 6.32 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 16.14, 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

$$\nu(f) = \operatorname{Re}\nu(f) = \operatorname{Re}\int_X f\rho d\lambda = \int_X f \operatorname{Re}\rho \, d\lambda. \tag{17.1}$$

Thus by replacing ρ by $\operatorname{Re}\rho$ if necessary we may assume ρ is real.

Taking $f = 1_{\rho<0}$ in Eq. (17.1) 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 17.2. *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 17.3 (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 17.4. *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 17.5. *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 $C \in \mathcal{M}$,

$$\nu(C \cap A) = \nu_s(C \cap A) + \nu_a(C \cap A) = \nu_s(C \cap A) = \nu_s(C) \tag{17.2}$$

and

$$\nu(C \cap A^c) = \nu_s(C \cap A^c) + \nu_a(C \cap A^c) = \nu_a(C \cap A^c) = \nu_a(C). \tag{17.3}$$

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. (17.2) and (17.3) with A being replaced by B to conclude,

$$\begin{aligned} \nu_s(C) &= \nu(C \cap B) = \tilde{\nu}_s(C) \text{ and} \\ \nu_a(C) &= \nu(C \cap B^c) = \tilde{\nu}_a(C) \text{ for all } C \in \mathcal{M}. \end{aligned}$$

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 17.6. *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,*

$$\int_A f d\mu = \int_A g d\mu \text{ for all } A \in \mathcal{M}. \tag{17.4}$$

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. (17.4) 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 17.7. Lemma 17.6 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. (17.4) holds yet $f \neq g$.

Theorem 17.8 (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 17.5 and 17.6.

Existence when μ and ν are both finite measures. (Von-Neumann's Proof. See Remark 17.9 for the motivation for this proof.) First suppose that μ and ν are **finite** measures and let $\lambda = \mu + \nu$. By Theorem 17.1, $d\nu = hd\lambda$

with $0 \le h \le 1$ and this implies, for all non-negative measurable functions f, that

$$\nu(f) = \lambda(fh) = \mu(fh) + \nu(fh) \tag{17.5}$$

or equivalently

$$\nu(f(1-h)) = \mu(fh). \tag{17.6}$$

Taking $f = 1_{\{h=1\}}$ in Eq. (17.6) shows that

$$\mu(\{h=1\}) = \nu(1_{\{h=1\}}(1-h)) = 0,$$

i.e. $0 \le h(x) < 1$ 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 $g \ge 0$ in Eq. (17.6) 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

$$\nu_n^s(X \setminus A_n) = \nu_n^s(X_n \setminus A_n) = 0.$$

This shows that $\nu_n^s \perp \mu$ for all n and so by Lemma 17.4, $\nu_s := \sum_{n=1}^{\infty} \nu_n^s$ is singular relative to μ. Since

$$\nu = \sum_{n=1}^{\infty} \nu_n = \sum_{n=1}^{\infty} (\rho_n \mu_n + \nu_n^s) = \sum_{n=1}^{\infty} (\rho_n 1_{X_n} \mu + \nu_n^s) = \rho\mu + \nu_s, \tag{17.7}$$

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 17.9. 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

$$\nu(1_A f) = \nu_s(1_A f) = \nu(1_A h f)$$

which show that $h = 1$, ν –a.e. on A. Also letting $f \to 1_B f$ implies that

$$\mu(\rho 1_B f) = \nu(h 1_B f) + \mu(h 1_B f) = \mu(\rho h 1_B f) + \mu(h 1_B f)$$

which implies, $\rho = \rho h + h$, μ − a.e. on B, i.e.

$$\rho\,(1 - h) = h, \ \mu\text{– a.e. on } B.$$

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,

$$d\nu = 1_{\{h=1\}} d\nu + 1_{\{h<1\}} d\nu = 1_{\{h=1\}} d\nu + \frac{h}{1-h} 1_{h<1} d\mu.$$

18

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 6.33 in this section. Because of this let us repeat it here.

Lemma 18.1. *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 $H = \sum a_i 1_{A_i}$ 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 $\bar{\mathbb{R}}$. ■

Lemma 18.2 (Integral Comparison). *Suppose that* $F, G : \Omega \to [0, \infty]$ *are* $\mathcal{B}$ *– measurable functions. Then* $F \geq G$ *a.s. iff*

$$\mathbb{E}[F : A] \geq \mathbb{E}[G : A] \text{ for all } A \in \mathcal{B}. \tag{18.1}$$

In particular $F = G$ *a.s. iff equality holds in Eq. (18.1). Moreover, for* $F \in L^1(\Omega, \mathcal{B}, P)$, $F = 0$ *a.s. iff* $\mathbb{E}[F : A] = 0$ *for all* $A \in \mathcal{B}$.

Proof. It is clear that $F \geq G$ a.s. implies Eq. (18.1). For the converse assertion, if we take $A = \{F = 0\}$ in Eq. (18.1) we learn that

$$0 = \mathbb{E}[F : F = 0] \geq \mathbb{E}[G : F = 0]$$

and hence that $G1_{F=0} = 0$ a.s., i.e.

$$G = 0 \text{ a.s. on } \{F = 0\}. \tag{18.2}$$

Similarly if $A := \{G > \alpha F\}$ with $\alpha > 1$ in Eq. (18.1), then

$$\mathbb{E}[F : G > \alpha F] \ge \mathbb{E}[G : G > \alpha F] \ge \mathbb{E}[\alpha F : G > \alpha F] = \alpha \mathbb{E}[F : G > \alpha F].$$

Since $\alpha > 1$, the only way this can happen is if $\mathbb{E}[F : G > \alpha F] = 0$. By the MCT we may now let $\alpha \downarrow 1$ to conclude, $0 = \mathbb{E}[F : G > F]$. This implies $F1_{G>F} = 0$ a.s. or equivalently

$$G \le F \text{ a.s on } \{F > 0\}. \tag{18.3}$$

Since $\Omega = \{F = 0\} \cup \{F > 0\}$ and on both sets, by Eqs. (18.2) and (18.3) we have $G \le F$ a.s. we may conclude that $G \le F$ a.s. on Ω as well. If equality holds in Eq. (18.1), then we know that $G \le F$ and $F \le G$ a.s., i.e. $F = G$ a.s.

If $F \in L^1(\Omega, \mathcal{B}, P)$ and $\mathbb{E}[F : A] = 0$ for all $A \in \mathcal{B}$, we may conclude by a simple limiting argument that $\mathbb{E}[Fh] = 0$ for all $h \in \mathcal{B}_b$. Taking $h := \operatorname{sgn}(F) := \frac{\bar{F}}{|F|} 1_{|F|>0}$ in this identity then implies

$$0 = \mathbb{E}[Fh] = \mathbb{E}\left[F \frac{\bar{F}}{|F|} 1_{|F|>0}\right] = \mathbb{E}\left[|F| \, 1_{|F|>0}\right] = \mathbb{E}[|F|]$$

which implies that $F = 0$ a.s. ■

Definition 18.3 (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 18.4 (Basic Properties of $\mathbb{E}_{\mathcal{G}}$). Let $f \in L^2(\Omega, \mathcal{B}, P)$. By the orthogonal projection Theorem 16.12 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,

$$\mathbb{E}_{\mathcal{G}_0} \mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1} \mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^2(\Omega, \mathcal{B}, P). \tag{18.4}$$

It is also useful to observe that condition 2. above may expressed as

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G} \tag{18.5}$$

or

$$\mathbb{E}[fh] = \mathbb{E}[Fh] \text{ for all } h \in \mathcal{G}_b. \tag{18.6}$$

Indeed, if Eq. (18.5) 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 6.32 and the DCT for all $h \in \mathcal{G}_b$. Therefore Eq. (18.5) implies Eq. (18.6). If Eq. (18.6) holds and $h \in L^2(\Omega, \mathcal{G}, P)$, we may use DCT to show

$$\mathbb{E}[fh] \overset{\text{DCT}}{=} \lim_{n\to\infty} \mathbb{E}\left[fh1_{|h|\le n}\right] \overset{(18.6)}{=} \lim_{n\to\infty} \mathbb{E}\left[Fh1_{|h|\le n}\right] \overset{\text{DCT}}{=} \mathbb{E}[Fh],$$

which is condition 2. in Remark 18.4. Taking $h = 1_A$ with $A \in \mathcal{G}$ in condition 2. or Remark 18.4, we learn that Eq. (18.5) is satisfied as well.

Theorem 18.5. *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 $f \ge 0$, P – a.e. then $\mathbb{E}_\mathcal{G} f \ge 0$, 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*
$$\mathbb{E}(Fh) = \mathbb{E}(fh) \text{ for all } h \in \mathcal{G}_b. \tag{18.7}$$
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*
$$\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1} f = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_0} f = \mathbb{E}_{\mathcal{G}_0} f \text{ a.s. for all } f \in L^1(\Omega, \mathcal{B}, P). \tag{18.8}$$

Proof. By the definition of orthogonal projection, $f \in L^2(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$,

$$\mathbb{E}(fh) = \mathbb{E}(f \cdot \mathbb{E}_\mathcal{G} h) = \mathbb{E}(\mathbb{E}_\mathcal{G} f \cdot h). \tag{18.9}$$

Taking

$$h = \overline{\operatorname{sgn}(\mathbb{E}_\mathcal{G} f)} := \frac{\overline{\mathbb{E}_\mathcal{G} f}}{\mathbb{E}_\mathcal{G} f} 1_{|\mathbb{E}_\mathcal{G} f| > 0} \tag{18.10}$$

in Eq. (18.9) shows

$$\mathbb{E}(|\mathbb{E}_\mathcal{G} f|) = \mathbb{E}(\mathbb{E}_\mathcal{G} f \cdot h) = \mathbb{E}(fh) \le \mathbb{E}(|fh|) \le \mathbb{E}(|f|). \tag{18.11}$$

It follows from this equation and the BLT (Theorem 16.17) 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. (18.9) remains valid for all $f \in L^1(\Omega, \mathcal{B}, P)$ and $h \in \mathcal{G}_b$. Indeed, (with out reference to Theorem 16.17) if $f_n := f1_{|f|\le n} \in L^2(\Omega, \mathcal{B}, P)$, 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 completness 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,

$$\mathbb{E}(F\cdot h)=\mathbb{E}(\lim_{n\to\infty}\mathbb{E}_{\mathcal{G}}f_n\cdot h)=\lim_{n\to\infty}\mathbb{E}(f_n\cdot h)=\mathbb{E}(f\cdot h) \tag{18.12}$$

for all $h\in\mathcal{G}_b$ and by Lemma 18.2 there is at most one, $F\in L^1(\Omega,\mathcal{G},P)$, which satisfies Eq. (18.12). 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. (18.7) of item 4. The same argument used in Eq. (18.11) 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

$$\mathbb{E}(\mathbb{E}_{\mathcal{G}}f\cdot h)=\mathbb{E}(fh)\ge 0\ \forall\ h\in\mathcal{G}_b \text{ with } h\ge 0. \tag{18.13}$$

An application of Lemma 18.2 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 repalced 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

$$\begin{aligned}\mathbb{E}\left[|\mathbb{E}_{\mathcal{G}}f|\,h\right]&=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}f\cdot\overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}}f)}h\right]=\mathbb{E}\left[f\cdot\overline{\operatorname{sgn}(\mathbb{E}_{\mathcal{G}}f)}h\right]\\&\le\mathbb{E}\left[|f|\,h\right]=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}|f|\cdot h\right].\end{aligned}$$

Since $h\ge 0$ is an arbitrary $\mathcal{G}$ – measurable function, it follows, by Lemma 18.2, 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

$$\mathbb{E}\left[(g\mathbb{E}_{\mathcal{G}}f)\,h\right]=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}f\cdot hg\right]=\mathbb{E}\left[f\cdot hg\right]=\mathbb{E}\left[gf\cdot h\right]=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}}(gf)\cdot h\right].$$

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 16.12, Eq. (18.8) 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. (18.8) continues to hold on $L^1(\Omega,\mathcal{B},P)$.

Second Proof. For $h\in(\mathcal{G}_0)_b$, we have

$$\mathbb{E}\left[\mathbb{E}_{\mathcal{G}_0}\mathbb{E}_{\mathcal{G}_1}f\cdot h\right]=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}_1}f\cdot h\right]=\mathbb{E}\left[f\cdot h\right]=\mathbb{E}\left[\mathbb{E}_{\mathcal{G}_0}f\cdot h\right]$$

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

$$\mathbb{E}_{\mathcal{G}_1}\left[\mathbb{E}_{\mathcal{G}_0}f\right]=\mathbb{E}_{\mathcal{G}_1}\left[\mathbb{E}_{\mathcal{G}_0}f\cdot 1\right]=\mathbb{E}_{\mathcal{G}_0}f\cdot\mathbb{E}_{\mathcal{G}_1}[1]=\mathbb{E}_{\mathcal{G}_0}f \text{ a.s.}$$

Notice that $\mathbb{E}_{\mathcal{G}_1}\left[\mathbb{E}_{\mathcal{G}_0}f\right]$ need only be $\mathcal{G}_1$ – measurable. What the statement says there are representatives of $\mathbb{E}_{\mathcal{G}_1}\left[\mathbb{E}_{\mathcal{G}_0}f\right]$ which is $\mathcal{G}_0$ – measurable and any such representative is also a representative of $\mathbb{E}_{\mathcal{G}_0}f$. ■

[1] This can also easily be proved directly here by taking $h=1_{\mathbb{E}_{\mathcal{G}}f<0}$ in Eq. (18.13).

Remark 18.6. There is another standard construction of $\mathbb{E}_{\mathcal{G}} f$ based on the characterization in Eq. (18.7) and the Radon Nikodym Theorem 17.8. It goes as follows, for $0 \leq f \in L^1(P)$, let $Q := fP$ and observe that $Q|_{\mathcal{G}} \ll P|_{\mathcal{G}}$ and hence there exists $0 \leq g \in L^1(\Omega, \mathcal{G}, P)$ 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 18.7 *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 18.1. 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 $-\infty \leq a < b \leq \infty$, 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.

18.1 Examples

Example 18.8. Suppose $\mathcal{G}$ is the trivial σ – algebra, i.e. $\mathcal{G} = \{\emptyset, \Omega\}$. In this case $\mathbb{E}_{\mathcal{G}} f = \mathbb{E} f$ a.s.

Example 18.9. On the opposite extreme, if $\mathcal{G} = \mathcal{B}$, then $\mathbb{E}_{\mathcal{G}} f = f$ a.s.

Lemma 18.10. *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 18.11 (**Note well.**).* According to Lemma 18.1, $\mathbb{E}(f|X) = \tilde{f}(X)$ 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

$$\mathbb{E}[f \cdot h(X)] = \mathbb{E}\left[\tilde{f}(X)\, h(X)\right] \tag{18.14}$$

for all bounded and measurable functions, $h : \mathbb{X} \to \mathbb{R}$.

Exercise 18.2. 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 σ – 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
$$\mathbb{E}_{\mathcal{G}} f = \sum_{i=1}^{\infty} \mathbb{E}[f|A_i]\, 1_{A_i} \text{ a.s.} \tag{18.15}$$

Solution to Exercise (18.2). 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. (18.15).

Proposition 18.12. *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 $0 \leq \rho \in L^1(\Omega, \mathcal{B}, \mu \otimes \nu)$ 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*

$$\bar{\rho}(x) := \int_{\mathbb{Y}} \rho(x, y)\, d\nu(y) \tag{18.16}$$

and $x \in \mathbb{X}$ and $B \in \mathcal{N}$, let

$$Q(x,B) := \begin{cases} \frac{1}{\bar{\rho}(x)} \int_B \rho(x,y)\, d\nu(y) & \text{if } \bar{\rho}(x) \in (0,\infty) \\ \delta_{y_0}(B) & \text{if } \bar{\rho}(x) \in \{0,\infty\} \end{cases} \tag{18.17}$$

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

$$\mathbb{E}[f(X,Y)|X] = Q(X, f(X,\cdot)) =: \int_{\mathbb{Y}} f(X,y)\, Q(X,dy) = g(X) \quad a.s. \tag{18.18}$$

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 18.11, we are searching for a bounded measurable function, $g : \mathbb{X} \to \mathbb{R}$, such that

$$\mathbb{E}[f(X,Y)\, h(X)] = \mathbb{E}[g(X)\, h(X)] \text{ for all } h \in \mathcal{M}_b. \tag{18.19}$$

(Throughout this argument we are going to repeatedly use the Tonelli - Fubini theorems.) We now explicitly write out both sides of Eq. (18.19);

$$\begin{aligned} \mathbb{E}[f(X,Y)\, h(X)] &= \int_{\mathbb{X}\times\mathbb{Y}} h(x)\, f(x,y)\, \rho(x,y)\, d\mu(x)\, d\nu(y) \\ &= \int_{\mathbb{X}} h(x) \left[\int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y)\right] d\mu(x) \end{aligned} \tag{18.20}$$

$$\begin{aligned} \mathbb{E}[g(X)\, h(X)] &= \int_{\mathbb{X}\times\mathbb{Y}} h(x)\, g(x)\, \rho(x,y)\, d\mu(x)\, d\nu(y) \\ &= \int_{\mathbb{X}} h(x)\, g(x)\, \bar{\rho}(x)\, d\mu(x). \end{aligned} \tag{18.21}$$

Since the right sides of Eqs. (18.20) and (18.21) must be equal for all $h \in \mathcal{M}_b$, we must demand,

$$\int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y) = g(x)\, \bar{\rho}(x) \text{ for } \mu - \text{a.e. } x. \tag{18.22}$$

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 μ – 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 ν – a.e. y and therefore

$$\int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y) = 0. \tag{18.23}$$

Hence at such points, x where $\bar{\rho}(x) = 0$, Eq. (18.22) 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 where 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. (18.21) we have

$$\begin{aligned}
\mathbb{E}[g(X)\, h(X)] &= \int_{\mathbb{X}} h(x)\, g(x)\, \bar{\rho}(x)\, d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\, g(x)\, \bar{\rho}(x)\, d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x)\, \bar{\rho}(x) \left(\frac{1}{\bar{\rho}(x)} \int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y) \right) d\mu(x) \\
&= \int_{\mathbb{X}\cap\{0<\bar{\rho}<\infty\}} h(x) \left(\int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y) \right) d\mu(x) \\
&= \int_{\mathbb{X}} h(x) \left(\int_{\mathbb{Y}} f(x,y)\, \rho(x,y)\, d\nu(y) \right) d\mu(x) \\
&= \mathbb{E}[f(X,Y)\, h(X)] \qquad \text{(by Eq. (18.20))},
\end{aligned}$$

wherein we have repeatedly used $\mu(\bar{\rho} = \infty) = 0$ and Eq. (18.23) 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 18.13. *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 18.14. *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. (18.17) 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

$$P_{\mathcal{G}}(A) = \sum_{n=1}^{\infty} P_{\mathcal{G}}(A_n) \text{ (equality in } L^2(\Omega, \mathcal{G}, P). \tag{18.24}$$

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. (18.24) it follows that

$$\bar{P}_{\mathcal{G}}(A)(\omega) = \sum_{n=1}^{\infty} \bar{P}_{\mathcal{G}}(A_n)(\omega) \text{ for } P\text{-a.e. } \omega. \tag{18.25}$$

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. (18.25) holds for all $\omega \in \Omega$ and all $A, A_n \in \mathcal{B}$ with $A = \sum_{n=1}^{\infty} A_n$.

Remark 18.15. 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 18.25. Moreover, it is known that all "reasonable" measure spaces are standard Borel spaces, see Section 18.4 below for more details. So in most instances of interest a regular conditional distribution of Y given X **will** exist.

Exercise 18.3. 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
$$Q(X, f(X,\cdot)) = \mathbb{E}[f(X,Y)|X] \text{ a.s.} \tag{18.26}$$
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
$$P((X,Y)\in A) = \int_{\mathbb{X}} Q(x, 1_A(x,\cdot))\, d\mu(x) = \int_{\mathbb{X}} d\mu(x) \int_{\mathbb{Y}} 1_A(x,y)\, Q(x,dy). \tag{18.27}$$

Exercise 18.4. Keeping the same notation as in Exercise 18.3 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.

$$\mathbb{E}[f(X,Y)|X] = \mathbb{E}[f(x,Y)|X]|_{x=X} \text{ a.s.}$$

Exercise 18.5. 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. $(X,Y) \stackrel{d}{=} (X',Y')$. 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

$$\mathbb{E}'[f(X',Y')|X'] = \tilde{f}(X') \ \ P' \text{ a.s.}$$

18.2 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 18.16 (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*

$$\mathbb{E}[f : A] = \mathbb{E}[F : A] \text{ for all } A \in \mathcal{G}. \tag{18.28}$$

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 18.5 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 $h \ge 0$ 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, $0 \le f_n \le f_{n+1}$ 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 $0 \le f_n \in L^1(\Omega, \mathcal{B}, P)$ a.s., then*

$$\mathbb{E}_{\mathcal{G}}\left[\liminf_{n\to\infty} f_n\right] \le \liminf_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f_n] \ \text{ a.s.} \tag{18.29}$$

4. ***Conditional Dominated Convergence (cDCT).*** *If $f_n \to f$ a.s. and $|f_n| \le g \in L^1(\Omega, \mathcal{B}, P)$, then $\mathbb{E}_{\mathcal{G}} f_n \to \mathbb{E}_{\mathcal{G}} f$ a.s.*

Remark 18.17. Regarding item 4. above. Suppose that $f_n \overset{P}{\to} f$, $|f_n| \le g_n \in L^1(\Omega, \mathcal{B}, P)$, $g_n \overset{P}{\to} g \in L^1(\Omega, \mathcal{B}, P)$ and $\mathbb{E}g_n \to \mathbb{E}g$. Then by the DCT in Corollary 11.8, 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. (18.28) holds and this uniquely determines F follows from Lemma 18.2.

Item 1. a) If $0 \le f \le g$, then

$$\mathbb{E}_{\mathcal{G}} f = \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[f \wedge n] \le \lim_{n\to\infty} \mathbb{E}_{\mathcal{G}}[g \wedge n] = \mathbb{E}_{\mathcal{G}} g \text{ a.s.}$$

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, $0 \le f_n \le f_{n+1}$ 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 18.19 shows how to localize conditional expectations. We first need the following definition.

Definition 18.18. *Suppose that $\mathcal{F}$ and $\mathcal{G}$ are sub-sigma-fileds 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. Indeed, if $B \in \mathcal{F}_A$ then $B \in \mathcal{G}_A$ and so $B \cap A \in \mathcal{F} \cap \mathcal{G}$ and hence $B \cap A = (B \cap A) \cap A \in [\mathcal{F} \cap \mathcal{G}]_A$. Moreover, because $[\mathcal{F} \cap \mathcal{G}]_A \subset \mathcal{F}_A$ we have $\mathcal{F}_A = \mathcal{G}_A$ implies

$$\mathcal{F}_A = \mathcal{G}_A = [\mathcal{F} \cap \mathcal{G}]_A . \tag{18.30}$$

Lemma 18.19 (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*

$$\mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F} \cap \mathcal{G}} X = \mathbb{E}_{\mathcal{F} \cap \mathcal{G}} Y = \mathbb{E}_{\mathcal{G}} Y \text{ a.s. on } A. \tag{18.31}$$

Alternatively put, if $A \in \mathcal{F} \cap \mathcal{G}$ and $\mathcal{F}_A = \mathcal{G}_A$ then

$$1_A \mathbb{E}_{\mathcal{F}} = 1_A \mathbb{E}_{\mathcal{F} \cap \mathcal{G}} = 1_A \mathbb{E}_{\mathcal{G}}. \tag{18.32}$$

Proof. Let us start with the observation that if X is an $\mathcal{F}$ – measurable random variable, then $1_A X$ is $\mathcal{F} \cap \mathcal{G}$ measurable. This can be checked directly (see Remark 18.20 below) or as follows. If $X = 1_B$ with $B \in \mathcal{F}$, then $1_A 1_B = 1_{A \cap B}$ 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 $1_A 1_B$ 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 $X \geq 0$ 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. (18.32) while the second follows by interchanging the roles of $\mathcal{F}$ and $\mathcal{G}$.

Equation (18.31) is now easily verified. First notice that $X = Y$ a.s. on A iff $1_A X = 1_A Y$ a.s.. Now from Eq. (18.32), the tower property of conditional expectation, and the fact that $1_A = 1_A \cdot 1_A$,we find

$$1_A \mathbb{E}_{\mathcal{F}} X = 1_A \mathbb{E}_{\mathcal{F}} \left[1_A X \right] = 1_A \mathbb{E}_{\mathcal{F}} \left[1_A Y \right] = 1_A \mathbb{E}_{\mathcal{F}} Y = 1_A \mathbb{E}_{\mathcal{F} \cap \mathcal{G}} Y$$

from which it follows that $\mathbb{E}_{\mathcal{F}} X = \mathbb{E}_{\mathcal{F} \cap \mathcal{G}} Y$ a.s. on A. ■

Remark 18.20. For the direct verification that $1_A X$ is $\mathcal{F} \cap \mathcal{G}$ measurable, we have,

$$\{1_A X \neq 0\} = A \cap \{X \neq 0\} \in \mathcal{F}_A = \mathcal{G}_A = (\mathcal{F} \cap \mathcal{G})_A \subset \mathcal{F} \cap \mathcal{G}.$$

So for $B \in \mathcal{B}_{\mathbb{R}}$,

$$\{1_A X \in B\} = A \cap \{X \in B\} \in \mathcal{F}_A \subset \mathcal{F} \cap \mathcal{G} \text{ if } 0 \notin B$$

while if $0 \in B$,

$$\begin{aligned} \{1_A X \in B\} &= \{1_A X = 0\}^c \cup A \cap \{X \in (B \setminus \{0\})\} \\ &= \{1_A X \neq 0\}^c \cup A \cap \{X \in (B \setminus \{0\})\} \in \mathcal{F} \cap \mathcal{G}. \end{aligned}$$

Theorem 18.21 (Conditional Jensen's inequality). *Let* $(\Omega, \mathcal{B}, P)$ *be a probability space,* $-\infty \leq a < b \leq \infty$, *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* $\varphi(\mathbb{E}_{\mathcal{G}} f) \in L^1(\Omega, \mathcal{G}, P)$,

$$\varphi(\mathbb{E}_{\mathcal{G}} f) \leq \mathbb{E}_{\mathcal{G}}[\varphi(f)] \ \text{ a.s.} \tag{18.33}$$

and

$$\mathbb{E}[\varphi(\mathbb{E}_{\mathcal{G}} f)] \leq \mathbb{E}[\varphi(f)] \tag{18.34}$$

Proof. Let $\Lambda := \mathbb{Q} \cap (a,b)$ – a countable dense subset of (a,b). By Theorem 11.38 (also see Lemma 7.31) and Figure 7.2 when φ is C^1)

$$\varphi(y) \geq \varphi(x) + \varphi'_-(x)(y-x) \text{ for all for all } x, y \in (a,b),$$

where $\varphi'_-(x)$ is the left hand derivative of φ at x. Taking $y = f$ and then taking conditional expectations imply,

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)] \geq \mathbb{E}_{\mathcal{G}}\left[\varphi(x) + \varphi'_-(x)(f-x)\right] = \varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x) \text{ a.s.} \tag{18.35}$$

Since this is true for all $x \in (a,b)$ (and hence all x in the countable set, Λ) we may conclude that

$$\mathbb{E}_{\mathcal{G}}[\varphi(f)] \geq \sup_{x \in \Lambda}\left[\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x)\right] \text{ a.s.}$$

By Exercise 18.1, $\mathbb{E}_{\mathcal{G}} f \in (a,b)$, and hence it follows from Corollary 11.39 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. (18.33).

From Eqs. (18.33) and (18.35) we infer,

$$|\varphi(\mathbb{E}_{\mathcal{G}} f)| \leq |\mathbb{E}_{\mathcal{G}}[\varphi(f)]| \vee \left|\varphi(x) + \varphi'_-(x)(\mathbb{E}_{\mathcal{G}} f - x)\right| \in L^1(\Omega, \mathcal{G}, P)$$

and hence $\varphi(\mathbb{E}_{\mathcal{G}} f) \in L^1(\Omega, \mathcal{G}, P)$. Taking expectations of Eq. (18.33) is now allowed and immediately gives Eq. (18.34). ∎

Corollary 18.22. *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* $1 \leq p \leq \infty$.

Proof. The case $p = \infty$ and $p = 1$ have already been covered in Theorem 18.5. So now suppose, $1 < p < \infty$, and apply Jensen's inequality with $\varphi(x) = |x|^p$ to find $|\mathbb{E}_{\mathcal{G}} f|^p \leq \mathbb{E}_{\mathcal{G}} |f|^p$ a.s. Taking expectations of this inequality gives the desired result. ∎

18.3 Regular Conditional Distributions

Lemma 18.23. *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 18.24. *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 exits 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 18.23 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

$$\begin{aligned}\mathbb{E}\left[1_{Y\leq r}\cdot g\left(X\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[1_{Y\leq r}|X\right]g\left(X\right)\right] = \mathbb{E}\left[q_r\left(X\right)g\left(X\right)\right] \\ &= \mathbb{E}\left[q_r\left(X\right)1_{\mathbb{X}_0}\left(X\right)g\left(X\right)\right].\end{aligned}$$

For $t\in\mathbb{R}$, we may let $r\downarrow t$ in the above equality (use DCT) to learn,

$$\mathbb{E}\left[1_{Y\leq t}\cdot g\left(X\right)\right] = \mathbb{E}\left[F\left(X,t\right)1_{\mathbb{X}_0}\left(X\right)g\left(X\right)\right] = \mathbb{E}\left[F\left(X,t\right)g\left(X\right)\right].$$

Since g was arbitrary, we may conclude that

$$Q\left(X,1_{(-\infty,t]}\right) = F\left(X,t\right) = \mathbb{E}\left[1_{Y\leq t}|X\right] \text{ a.s.}$$

This completes the proof. ■

This result leads fairly immediately to the following far reaching generalization.

Theorem 18.25. *Suppose that $(\mathbb{X},\mathcal{M})$ is a measurable space and $(\mathbb{Y},\mathcal{N})$ is a standard Borel space, see Appendix 18.4 below. Suppose that $X:\Omega\to\mathbb{X}$ and $Y:\Omega\to\mathbb{Y}$ are measurable functions. Then there exits a probability kernel, Q, on $\mathbb{X}\times\mathbb{Y}$ such that $\mathbb{E}\left[f\left(Y\right)|X\right] = Q\left(X,f\right)$, 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 $Y\left(\Omega\right)\subset\mathbb{Y}$. By Theorem 18.24, we may find a probability kernel, Q_0, on $\mathbb{X}\times\mathbb{R}$ such that

$$\mathbb{E}\left[f\left(Y\right)|X\right] = Q_0\left(X,f\right),\ P-a.s., \tag{18.36}$$

for all bounded measurable functions, $f:\mathbb{R}\to\mathbb{R}$.

Taking $f=1_{\mathbb{Y}}$ in Eq. (18.36) shows

$$1 = \mathbb{E}\left[1_{\mathbb{Y}}\left(Y\right)|X\right] = Q_0\left(X,\mathbb{Y}\right) \text{ a.s..}$$

Thus if we let $\mathbb{X}_0 := \{x\in\mathbb{X}: Q_0\left(x,\mathbb{Y}\right)=1\}$, we know that $P\left(X\in\mathbb{X}_0\right)=1$. Let us now define

$$Q\left(x,B\right) := 1_{\mathbb{X}_0}\left(x\right)Q_0\left(x,B\right) + 1_{\mathbb{X}\setminus\mathbb{X}_0}\left(x\right)\delta_y\left(B\right) \text{ for } \left(x,B\right)\in\mathbb{X}\times\mathcal{B}_{\mathbb{Y}},$$

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

$$Q\left(X,B\right) = 1_{\mathbb{X}_0}\left(X\right)Q_0\left(X,B\right) = 1_{\mathbb{X}_0}\left(X\right)\mathbb{E}\left[1_B\left(Y\right)|X\right] = \mathbb{E}\left[1_B\left(Y\right)|X\right] \text{ a.s.}$$

This shows that Q is the desired regular conditional probability. ■

Corollary 18.26. *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 exits a probability kernel, Q, on $(\Omega,\mathcal{G})\times(\mathbb{Y},\mathcal{N})$ such that $\mathbb{E}\left[f\left(Y\right)|\mathcal{G}\right] = Q\left(\cdot,f\right)$, P - a.s. for all bounded measurable functions, $f:\mathbb{Y}\to\mathbb{R}$.*

Proof. This is a special case of Theorem 18.25 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. ■

18.4 Appendix: Standard Borel Spaces

For more information along the lines of this section, see Royden [**?**].

Definition 18.27. *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 18.28. *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 18.29 (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 18.39 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 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 18.36.

On first reading, you may wish to skip the rest of this section.

Lemma 18.30. *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 18.31. *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 18.30,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 18.30 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 18.30 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 18.32 *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 18.33. *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 18.34. *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 $r_1(x) = x$,

$$\gamma_1(x) := 1_{x \geq 2^{-1}} \text{ and } r_2(x) := x - 2^{-1}\gamma_1(x) \in (0, 2^{-1}),$$

then let $\gamma_2 := 1_{r_2 \geq 2^{-2}}$ 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

$$r_{k+1}(x) = r_k(x) - 2^{-k}\gamma_k(x) = x - \sum_{j=1}^{k} 2^{-j}\gamma_j(x) \in \left(0, 2^{-k}\right) \qquad (18.37)$$

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

$$x = \sum_{j=1}^{\infty} 2^{-j}\gamma_j(x) \text{ and } r_{k+1}(x) = \sum_{j=k+1}^{\infty} 2^{-j}\gamma_j(x). \qquad (18.38)$$

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 $j \geq k$, then (by Eq. (18.38)) $r_{k+1}(x) = 2^{-k}$ which would contradict Eq. (18.37). 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 $x \geq 2^{-1}$ 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 $r_2(x) \geq 2^{-2}$ 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

$$\Lambda := \Omega \setminus \Omega_0 = \cup_{n=1}^{\infty} \{\omega \in \Omega : \omega_j = 1 \text{ for } j \geq n\}$$

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 18.30 and Proposition 18.31 now implies

$$\Omega = \Omega_0 \cup \Lambda \cong [0,1) \cup \mathbb{N} \cong [0,1) \cong (0,1).$$

■

Corollary 18.35. *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 18.30 and Proposition 18.31 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 18.34, 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 18.33. 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 18.36. *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 18.35, 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 18.37 *Let $Q := [0,1]^{\mathbb{N}}$ denote the (infinite dimensional)* ***unit cube*** *in $\mathbb{R}^{\mathbb{N}}$. For $a, b \in Q$ let*

$$d(a,b) := \sum_{n=1}^\infty \frac{1}{2^n} |a_n - b_n| = \sum_{n=1}^\infty \frac{1}{2^n} |\pi_n(a) - \pi_n(b)|. \tag{18.39}$$

Exercise 18.6. 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 (18.6). It is easily seen that d is a metric on Q which, by Eq. (18.39) 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

$$|\pi_n(a) - \pi_n(b)| \leq 2^n d(a, b),$$

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 18.38. *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 [**?**, Theorem 82.5 on p. 106.].) By replacing ρ by $\frac{\rho}{1+\rho}$ if necessary, we may assume that $0 \leq \rho < 1$. 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,

$$\begin{aligned}\rho(x, y) &\leq \rho(x, a_n) + \rho(y, a_n) = 2\rho(x, a_n) + \rho(y, a_n) - \rho(x, a_n) \\ &\leq 2\rho(x, a_n) + |\rho(x, a_n) - \rho(y, a_n)| \leq 2\rho(x, a_n) + 2^n \gamma(x, y).\end{aligned}$$

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 $N_1 \cap \dots \cap N_n$ is an open neighborhood of $u \in \bar{S}$, there exists $u_n \in N_1 \cap \dots \cap N_n \cap S$ 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 $S_{1/n} := \left\{u \in Q : d\left(u, \bar{S}\right) < 1/n\right\}$ 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 18.39. *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 18.38 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. ■

19

(Sub and Super) Martingales

Notation 19.1 *A* ***filtered probability space*** *is a probability space,* $(\Omega, \mathcal{B}, P)$ *endowed with a sequence of sub-*σ *-algebras,* $\{\mathcal{B}_n\}_{n=0}^\infty$ *such that* $\mathcal{B}_n \subset \mathcal{B}_{n+1} \subset \mathcal{B}$ *for all* $n = 0, 1, 2 \dots$. *We further define*

$$\mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma\left(\cup_{n=0}^\infty \mathcal{B}_n\right) \subset \mathcal{B}. \tag{19.1}$$

Through out this chapter, we will assume $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $\mathcal{B}_\infty$ is defined as in Eq. (19.1).

Definition 19.2. *A sequence of random variables,* $\{Y_n\}_{n=0}^\infty$ *are* ***adapted*** *to the filtration if* Y_n *is* $\mathcal{B}_n$ *– measurable for all* n. *We say* $\{Z_n\}_{n=1}^\infty$ *is* ***predictable*** *if each* Z_n *is* $\mathcal{B}_{n-1}$ *– measurable for all* $n \in \mathbb{N}$.

A typical example is when $\{X_n\}_{n=0}^\infty$ is a sequence of random variables on a probability space $(\Omega, \mathcal{B}, P)$ and $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$. An application of Lemma 18.1 shows that $\{Y_n\}$ is adapted to the filtration iff there are measurable functions, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $Y_n = f_n(X_0, \dots, X_n)$ for all $n \in \mathbb{N}_0$ and $\{Z_n\}_{n=1}^\infty$ is predictable iff there exists, there are measurable functions, $f_n : \mathbb{R}^n \to \mathbb{R}$ such that $Z_n = f_n(X_0, \dots, X_{n-1})$ for all $n \in \mathbb{N}$.

Definition 19.3. *Let* $X := \{X_n\}_{n=0}^\infty$ *is a be an adapted sequence of integrable random variables. Then;*

1. *X is a* $\{\mathcal{B}_n\}_{n=0}^\infty$ *–* ***martingale*** *if* $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ *a.s. for all* $n \in \mathbb{N}_0$.
2. *X is a* $\{\mathcal{B}_n\}_{n=0}^\infty$ *–* ***submartingale*** *if* $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \geq X_n$ *a.s. for all* $n \in \mathbb{N}_0$.
3. *X is a* $\{\mathcal{B}_n\}_{n=0}^\infty$ *–* ***supermartingale*** *if* $\mathbb{E}[X_{n+1}|\mathcal{B}_n] \leq X_n$ *a.s. for all* $n \in \mathbb{N}_0$.

By induction one shows that X is a supermartingale, martingale, or submartingale iff

$$\mathbb{E}[X_m|\mathcal{B}_n] \overset{\leq}{\underset{\geq}{=}} X_n \text{ a.s for all } m \geq n, \tag{19.2}$$

to be read from top to bottom respectively. This last equation may also be expressed as

$$\mathbb{E}[X_m|\mathcal{B}_n] \overset{\leq}{\underset{\geq}{=}} X_{m \wedge n} \text{ a.s for all } m, n \in \mathbb{N}_0. \tag{19.3}$$

The reader should also note that $\mathbb{E}[X_n]$ is decreasing, constant, or increasing respectively. The next lemma shows that we may shrink the filtration, $\{\mathcal{B}_n\}_{n=0}^\infty$, within limits and still have X retain the property of being a supermartingale, martingale, or submartingale.

Lemma 19.4 (Shrinking the filtration). *Suppose that X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively and $\{\mathcal{B}_n'\}_{n=0}^\infty$ is another filtration such that $\sigma(X_0,\dots,X_n) \subset \mathcal{B}_n' \subset \mathcal{B}_n$ for all n. Then X is a $\{\mathcal{B}_n'\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively.*

Proof. Since $\{X_n\}_{n=0}^\infty$ is adapted to $\{\mathcal{B}_n\}_{n=0}^\infty$ and $\sigma(X_0,\dots,X_n) \subset \mathcal{B}_n' \subset \mathcal{B}_n$, for all n,

$$\mathbb{E}_{\mathcal{B}_n'} X_{n+1} = \mathbb{E}_{\mathcal{B}_n'} \mathbb{E}_{\mathcal{B}_n} X_{n+1} \overset{\leq}{\underset{\geq}{=}} \mathbb{E}_{\mathcal{B}_n'} X_n = X_n,$$

when X is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale respectively – read from top to bottom. ■

Enlarging the filtration is another matter all together. In what follows we will simply say X is a supermartingale, martingale, submartingale if it is a $\{\mathcal{B}_n\}_{n=0}^\infty$ – supermartingale, martingale, submartingale.

19.1 (Sub and Super) Martingale Examples

Example 19.5. Suppose that $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and $X \in L^1(\Omega, \mathcal{B}, P)$. Then $X_n := \mathbb{E}[X|\mathcal{B}_n]$ is a martingale. Indeed, by the tower property of conditional expectations,

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_{n+1}]|\mathcal{B}_n] = \mathbb{E}[X|\mathcal{B}_n] = X_n \text{ a.s.}$$

Example 19.6. Suppose that $\Omega = [0,1]$, $\mathcal{B} = \mathcal{B}_{[0,1]}$, and $P = m$ – Lebesgue measure. Let $\mathcal{P}_n = \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right\}_{k=1}^{2^n-1} \cup \left\{\left[0, \frac{1}{2^n}\right]\right\}$ and $\mathcal{B}_n := \sigma(\mathcal{P}_n)$ for each $n \in \mathbb{N}$. Then $M_n := 2^n 1_{(0,2^{-n}]}$ for $n \in \mathbb{N}$ is a martingale such that $\mathbb{E}|M_n| = 1$ for all n. However, there is no $X \in L^1(\Omega, \mathcal{B}, P)$ such that $M_n = \mathbb{E}[X|\mathcal{B}_n]$. To verify this last assertion, suppose such an X existed. Let . We would then have for $2^n > k > 0$ and any $m > n$, that

$$\begin{aligned}\mathbb{E}\left[X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] &= \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_m} X : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] \\ &= \mathbb{E}\left[M_m : \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right] = 0.\end{aligned}$$

Using $\mathbb{E}[X : A] = 0$ for all A in the π – system, $\mathcal{Q} := \cup_{n=1}^\infty \left\{\left(\frac{k}{2^n}, \frac{k+1}{2^n}\right] : 0 < k < 2^n\right\}$, an application of the $\pi - \lambda$ theorem shows $\mathbb{E}[X : A] = 0$ for all $A \in \sigma(\mathcal{Q}) = \mathcal{B}$.

Therefore $X = 0$ a.s. by Lemma 18.2. But this is impossible since $1 = \mathbb{E}M_n = \mathbb{E}X$.

Moral: not all L^1 – bounded martingales are of the form in example 19.5. Proposition 19.7 shows what is missing from this martingale in order for it to be of the form in Example 19.5.

Proposition 19.7. *Suppose $1 \le p < \infty$ and $X \in L^p(\Omega, \mathcal{B}, P)$. Then the collection of random variables, $\Gamma := \{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subset \mathcal{B}\}$ is a bounded subset of $L^p(\Omega, \mathcal{B}, P)$ which is also uniformly integrable.*

Proof. Since $\mathbb{E}_\mathcal{G}$ is a contraction on all L^p – spaces it follows that Γ is bounded in L^p with

$$\sup_{\mathcal{G} \subset \mathcal{B}} \|\mathbb{E}[X|\mathcal{G}]\|_p \le \|X\|_p .$$

For the $p > 1$ the uniform integrability of Γ follows directly from Lemma 11.35.

We now concentrate on the $p = 1$ case. Recall that $|\mathbb{E}_\mathcal{G} X| \le \mathbb{E}_\mathcal{G} |X|$ a.s. and therefore,

$$\mathbb{E}[|\mathbb{E}_\mathcal{G} X| : |\mathbb{E}_\mathcal{G} X| \ge a] \le \mathbb{E}[|X| : |\mathbb{E}_\mathcal{G} X| \ge a] \text{ for all } a > 0.$$

But by Chebyshev's inequality,

$$P(|\mathbb{E}_\mathcal{G} X| \ge a) \le \frac{1}{a} \mathbb{E}|\mathbb{E}_\mathcal{G} X| \le \frac{1}{a} \mathbb{E}|X| .$$

Since $\{|X|\}$ is uniformly integrable, it follows from Proposition 11.29 that, by choosing a sufficiently large, $\mathbb{E}[|X| : |\mathbb{E}_\mathcal{G} X| \ge a]$ is as small as we please uniformly in $\mathcal{G} \subset \mathcal{B}$ and therefore,

$$\lim_{a \to \infty} \sup_{\mathcal{G} \subset \mathcal{B}} \mathbb{E}[|\mathbb{E}_\mathcal{G} X| : |\mathbb{E}_\mathcal{G} X| \ge a] = 0.$$

■

Example 19.8. This example generalizes Example 19.6. Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space and Q is another probability measure on $(\Omega, \mathcal{B})$. Let us assume that $Q|_{\mathcal{B}_n} \ll P|_{\mathcal{B}_n}$ for all n, which by the Raydon-Nikodym Theorem 17.8, implies there exists $0 \le X_n \in L^1(\Omega, \mathcal{B}_n, P)$ with $\mathbb{E}X_n = 1$ such that $dQ|_{\mathcal{B}_n} = X_n dP|_{\mathcal{B}_n}$, or equivalently put, for any $B \in \mathcal{B}_n$ we have

$$Q(B) = \int_B X_n dP = \mathbb{E}[X_n : B] .$$

Since $B \in \mathcal{B}_n \subset \mathcal{B}_{n+1}$, we also have $\mathbb{E}[X_{n+1} : B] = Q(B) = \mathbb{E}[X_n : B]$ for all $B \in \mathcal{B}_n$ and hence $\mathbb{E}[X_{n+1}|\mathcal{B}_n] = X_n$ a.s., i.e. $X = \{X_n\}_{n=0}^\infty$ is a positive martingale.

Example 19.6 is of this form with $Q = \delta_0$. Notice that $\delta_0|_{\mathcal{B}_n} \ll m|_{\mathcal{B}_n}$ for all $n < \infty$ while $\delta_0 \perp m$ on $\mathcal{B}_{[0,1]} = \mathcal{B}_\infty$. See Section **??** for more in the direction of this example.

It is often fruitful to view X_n as your earnings at time n while playing some game of chance. In this interpretation, your expected earnings at time $n+1$ given the history of the game up to time n is the same, greater than, less than your earnings at time n if $X = \{X_n\}_{n=0}^{\infty}$ is a martingale, submartingale or supermartingale respectively. In this interpretation, martingales are fair games, submartingales are favorable games, and supermartingales are unfavorable games.

Example 19.9. Suppose at each time n, we flip a fair coin and record the value, $X_n \in \{0,1\}$. Let us suppose that a gambler is going to bet one dollar between flips that either a 0 or a 1 is going to occur and if she is correct she will be paid $1+\alpha$ dollars in return, otherwise she loses his dollar to the house. Let us say Y_{n+1} is the gambler's prediction for the value of X_{n+1} at time n. Hence if we let M_n denote the gamblers fortune at time n we have

$$M_{n+1} = M_n - 1 + (1+\alpha)\, 1_{Y_{n+1}=X_{n+1}}.$$

Assuming the gambler can not see into the future, his/her prediction at time n can only depend on the game up to time n, i.e. we should have $Y_{n+1} = f_{n+1}(X_0, \dots, X_n)$ or equivalently, Y_{n+1} is $\mathcal{B}_n = \sigma(X_0, \dots, X_n)$ measurable. In this situation $\{M_n\}_{n=1}^{\infty}$ is an adapted process and moreover,

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{B}_n] &= \mathbb{E}\left[M_n - 1 + (1+\alpha)\, 1_{Y_{n+1}=X_{n+1}}|\mathcal{B}_n\right] \\
&= M_n - 1 + (1+\alpha)\, \mathbb{E}\left[1_{Y_{n+1}=X_{n+1}}|\mathcal{B}_n\right] \\
&= M_n - 1 + (1+\alpha)\, \frac{1}{2} = M_n + \frac{1}{2}(\alpha - 1)
\end{aligned}$$

wherein we have used Exercise 18.4 in the last line. Hence we see that $\{M_n\}_{n=1}^{\infty}$ is a martingale if $\alpha = 1$, a sub-martingale of $\alpha > 1$ and as supermartingale of $\alpha < 1$.

Exercise 19.1. Suppose that $\{X_n\}_{n=1}^{\infty}$ are i.i.d. random functions taking values in a finite set, S, and let $p(s) := P(X_n = s)$ for all $s \in S$ and assume $p(s) > 0$ for all s. As above let $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$ with $\mathcal{B}_0 = \{\emptyset, \Omega\}$ and suppose that $\alpha : S \to \mathbb{R}$ is a payoff function. Let Y_n be the predictions of a gambler as to the value of X_n based on the values of $\{X_1, \dots, X_{n-1}\}$, i.e. $Y_n \in S$ is a $\mathcal{B}_{n-1}$ – measurable random variable with the convention that $\mathcal{B}_0 = \{\emptyset, \Omega\}$. Also let M_n be the gambler's fortune at time n. Assuming the gambler always wages one dollar and receives a pay off of $1 + \alpha(s)$ if $Y_{n+1} = s = X_{n+1}$ for some $s \in S$, then

$$M_{n+1} = M_n - 1 + \sum_{s \in S} (1 + \alpha(s))\, 1_{Y_{n+1}=s=X_{n+1}}.$$

Show $\{M_n\}$ is a martingale, submartingale, supermartingale, if $\alpha = \frac{1-p}{p}$,[1] $\alpha \geq \frac{1-p}{p}$, or $\alpha \leq \frac{1-p}{p}$ respectively.

[1] In more detail, $\alpha(s) = \frac{1-p(s)}{p(s)}$ for all $s \in S$.

Lemma 19.10. *Let $X := \{X_n\}_{n=0}^{\infty}$ be an adapted process of integrable random variables on a filtered probability space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ and let $d_n := X_n - X_{n-1}$ with $X_{-1} := \mathbb{E}X_0$. Then X is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \geq 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \leq 0$ respectively) for all $n \in \mathbb{N}_0$.*

Conversely if $\{d_n\}_{n=1}^{\infty}$ is an adapted sequence of integrable random variables and X_0 is a $\mathcal{B}_0$ -measurable integral random variable. Then $X_n = X_0 + \sum_{j=1}^{n} d_j$ is a martingale (respectively submartingale or supermartingale) iff $\mathbb{E}[d_{n+1}|\mathcal{B}_n] = 0$ ($\mathbb{E}[d_{n+1}|\mathcal{B}_n] \geq 0$ or $\mathbb{E}[d_{n+1}|\mathcal{B}_n] \leq 0$ respectively) for all $n \in \mathbb{N}$.

Proof. We prove the assertions for martingales only, the other all being similar. Clearly X is a martingale iff

$$0 = \mathbb{E}[X_{n+1}|\mathcal{B}_n] - X_n = \mathbb{E}[X_{n+1} - X_n|\mathcal{B}_n] = \mathbb{E}[d_{n+1}|\mathcal{B}_n].$$

The second assertion is an easy consequence of the first assertion. ∎

Example 19.11. Suppose that $\{X_n\}_{n=0}^{\infty}$ is a sequence of independent random variables, $S_n = X_0 + \cdots + X_n$, and $\mathcal{B}_n := \sigma(X_0, \ldots, X_n) = \sigma(S_0, \ldots, S_n)$. Then

$$\mathbb{E}[S_{n+1}|\mathcal{B}_n] = \mathbb{E}[S_n + X_{n+1}|\mathcal{B}_n] = S_n + \mathbb{E}[X_{n+1}|\mathcal{B}_n] = S_n + \mathbb{E}[X_{n+1}].$$

Therefore $\{S_n\}_{n=0}^{\infty}$ is a martingale respectively submartingale or supermartingale) iff $\mathbb{E}X_n = 0$ ($\mathbb{E}X_n \geq 0$ or $\mathbb{E}X_n \leq 0$ respectively) for all $n \in \mathbb{N}$.

Example 19.12. Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a sequence of independent integrable random variables, $X_n = Z_0 \ldots Z_n$, and $\mathcal{B}_n := \sigma(Z_0, \ldots, Z_n)$. (Observe that $\mathbb{E}|X_n| = \prod_{k=0}^{n} \mathbb{E}|Z_k| < \infty$.) If $\mathbb{E}Z_n = 1$ for all n then X is a martingale while if $Z_n \geq 0$ and $\mathbb{E}Z_n \leq 1$ ($\mathbb{E}Z_n \geq 1$) for all n then X is a supermartingale (submartingale). Indeed, this follows from the simple identity;

$$\mathbb{E}[X_{n+1}|\mathcal{B}_n] = \mathbb{E}[X_n Z_{n+1}|\mathcal{B}_n] = X_n \mathbb{E}[Z_{n+1}|\mathcal{B}_n] = X_n \cdot \mathbb{E}[Z_{n+1}] \text{ a.s.}$$

Proposition 19.13. *Suppose that $X = \{X_n\}_{n=0}^{\infty}$ is a martingale and φ is a convex function such that $\varphi(X_n) \in L^1$ for all n. Then $\varphi(X) = \{\varphi(X_n)\}_{n=0}^{\infty}$ is a submartingale. If φ is also assumed to be increasing, it suffices to assume that X is a submartingale in order to conclude that $\varphi(X)$ is a submartingale. (For example if X is a positive submartingale, $p \in (1, \infty)$, and $\mathbb{E}X_n^p < \infty$ for all n, then $X^p := \{X_n^p\}_{n=0}^{\infty}$ is another positive submartingale.*

Proof. When X is a martingale, by the conditional Jensen's inequality 18.21,

$$\varphi(X_n) = \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \leq \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

which shows $\varphi(X)$ is a submartingale. Similarly, if X is a submartingale and φ is convex and increasing, then φ preserves the inequality, $X_n \leq \mathbb{E}_{\mathcal{B}_n} X_{n+1}$, and hence

$$\varphi(X_n) \leq \varphi(\mathbb{E}_{\mathcal{B}_n} X_{n+1}) \leq \mathbb{E}_{\mathcal{B}_n}[\varphi(X_{n+1})]$$

so again $\varphi(X)$ is a submartingale. ∎

19.2 Decompositions

Notation 19.14 *Given a sequence* $\{Z_k\}_{k=0}^{\infty}$, *let* $\Delta_k Z := Z_k - Z_{k-1}$ *for* $k = 1, 2, \dots$.

Lemma 19.15 (Doob Decomposition). *To each adapted sequence,* $\{Z_n\}_{n=0}^{\infty}$, *of integrable random variables has a unique decomposition,*

$$Z_n = M_n + A_n \tag{19.4}$$

where $\{M_n\}_{n=0}^{\infty}$ *is a martingale and* A_n *is a predictable process such that* $A_0 = 0$. *Moreover this decomposition is given by* $A_0 = 0$,

$$A_n := \sum_{k=1}^{n} \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z] \text{ for } n \geq 1 \tag{19.5}$$

and

$$M_n = Z_n - A_n = Z_n - \sum_{k=1}^{n} \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z] \tag{19.6}$$

$$= Z_0 + \sum_{k=1}^{n} \left(Z_k - \mathbb{E}_{\mathcal{B}_{k-1}} Z_k \right). \tag{19.7}$$

In particular, $\{Z_n\}_{n=0}^{\infty}$ *is a submartingale (supermartingale) iff* A_n *is increasing (decreasing) almost surely.*

Proof. Assuming Z_n has a decomposition as in Eq. (19.4), then

$$\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} Z] = \mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} M + \Delta_{n+1} A] = \Delta_{n+1} A \tag{19.8}$$

wherein we have used M is a martingale and A is predictable so that $\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} M] = 0$ and $\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} A] = \Delta_{n+1} A$. Hence we must define, for $m \geq 1$,

$$A_n := \sum_{k=1}^{n} \Delta_k A = \sum_{k=1}^{n} \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z]$$

which is a predictable process. This proves the uniqueness of the decomposition and the validity of Eq. (19.5).

For existence, from Eq. (19.5) it follows that

$$\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} Z] = \Delta_{n+1} A = \mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} A].$$

Hence, if we define $M_n := Z_n - A_n$, then

$$\mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} M] = \mathbb{E}_{\mathcal{B}_n} [\Delta_{n+1} Z - \Delta_{n+1} A] = 0$$

and hence $\{M_n\}_{n=0}^{\infty}$ is a martingale. Moreover, Eq. (19.7) follows from Eq. (19.6) since,

$$M_n = Z_0 + \sum_{k=1}^{n} \left(\Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z] \right)$$

and

$$\begin{aligned} \Delta_k Z - \mathbb{E}_{\mathcal{B}_{k-1}} [\Delta_k Z] &= Z_k - Z_{k-1} - \mathbb{E}_{\mathcal{B}_{k-1}} [Z_k - Z_{k-1}] \\ &= Z_k - Z_{k-1} - \left(\mathbb{E}_{\mathcal{B}_{k-1}} Z_k - Z_{k-1} \right) = Z_k - \mathbb{E}_{\mathcal{B}_{k-1}} Z_k. \end{aligned}$$

■

Remark 19.16. Suppose that $X = \{X_n\}_{n=0}^{\infty}$ is a submartingale and $X_n = M_n + A_n$ is it Doob decomposition. Then $A_\infty =\uparrow \lim_{n\to\infty} A_n$ exists a.s.,

$$\mathbb{E}A_n = \mathbb{E}[X_n - M_n] = \mathbb{E}X_n - \mathbb{E}M_0 = \mathbb{E}[X_n - X_0] \tag{19.9}$$

and hence by MCT,

$$\mathbb{E}A_\infty =\uparrow \lim_{n\to\infty} \mathbb{E}[X_n - X_0]. \tag{19.10}$$

Hence if $\lim_{n\to\infty} \mathbb{E}[X_n - X_0] = \sup_n \mathbb{E}[X_n - X_0] < \infty$, then $\mathbb{E}A_\infty < \infty$ and so by DCT, $A_n \to A_\infty$ in $L^1(\Omega, \mathcal{B}, P)$. In particular if $\sup_n \mathbb{E}|X_n| < \infty$, we may conclude that $\{X_n\}_{n=0}^{\infty}$ is $L^1(\Omega, \mathcal{B}, P)$ convergent iff $\{M_n\}_{n=0}^{\infty}$ is $L^1(\Omega, \mathcal{B}, P)$ convergent. (We will see below in Corollary 19.46 that $X_\infty := \lim_{n\to\infty} X_n$ and $M_\infty := \lim_{n\to\infty} M_n$ exist almost surely under the assumption that $\sup_n \mathbb{E}|X_n| < \infty$.)

Example 19.17. Suppose that $N = \{N_n\}_{n=0}^{\infty}$ is a square integrable martingale, i.e. $\mathbb{E}N_n^2 < \infty$ for all n. Then from Proposition 19.13, $X := \left\{X_n = N_n^2\right\}_{n=0}^{\infty}$ is a positive submartingale. In this case

$$\begin{aligned} \mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X &= \mathbb{E}_{\mathcal{B}_{k-1}} \left(N_k^2 - N_{k-1}^2 \right) = \mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1})(N_k + N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1})(N_k - N_{k-1})] \\ &= \mathbb{E}_{\mathcal{B}_{k-1}} (N_k - N_{k-1})^2 \end{aligned}$$

wherein the second to last equality we have used

$$\mathbb{E}_{\mathcal{B}_{k-1}} [(N_k - N_{k-1}) N_{k-1}] = N_{k-1} \mathbb{E}_{\mathcal{B}_{k-1}} (N_k - N_{k-1}) = 0 \text{ a.s.}$$

in order to change $(N_k + N_{k-1})$ to $(N_k - N_{k-1})$. Hence the increasing predictable process, A_n, in the Doob decomposition may be written as

$$A_n = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} \Delta_k X = \sum_{k\le n} \mathbb{E}_{\mathcal{B}_{k-1}} (\Delta_k N)^2. \tag{19.11}$$

For the next result we will use the following remarks.

Remark 19.18. If X is a real valued random variable, then $X = X^+ - X^-$, $|X| = X^+ + X^-$, $X^+ \le |X| = 2X^+ - X$, so that

$$\mathbb{E}X^+ \le \mathbb{E}|X| = 2\mathbb{E}X^+ - \mathbb{E}X.$$

Hence if $\{X_n\}_{n=0}^\infty$ is a submartingale then

$$\mathbb{E}X_n^+ \le \mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \le 2\mathbb{E}X_n^+ - \mathbb{E}X_0$$

from which it follows that

$$\sup_n \mathbb{E}X_n^+ \le \sup_n \mathbb{E}|X_n| \le 2\sup_n \mathbb{E}X_n^+ - \mathbb{E}X_0. \tag{19.12}$$

Theorem 19.19 (Krickeberg Decomposition). *Suppose that X is an integrable submartingale such that $C := \sup_n \mathbb{E}[X_n^+] < \infty$ or equivalently $\sup_n \mathbb{E}|X_n| < \infty$, see Eq. (19.12). Then*

$$M_n :=\uparrow \lim_{p\to\infty} \mathbb{E}\left[X_p^+ | \mathcal{B}_n\right] \text{ exists a.s.,}$$

$M = \{M_n\}_{n=0}^\infty$ is a positive martingale, $Y = \{Y_n\}_{n=0}^\infty$ with $Y_n := X_n - M_n$ is a positive supermartingale, and hence $X_n = M_n - Y_n$. So X can be decomposed into the difference of a positive martingale and a positive supermartingale.

Proof. From Proposition 19.13 we know that $X^+ = \{X_n^+\}$ is a still a positive submartingale. Therefore for each $n \in \mathbb{N}$, and $p \ge n$,

$$\mathbb{E}_{\mathcal{B}_n}\left[X_{p+1}^+\right] = \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_p}\left[X_{p+1}^+\right] \ge \mathbb{E}_{\mathcal{B}_n} X_p^+ \text{ a.s.}$$

Therefore $\mathbb{E}_{\mathcal{B}_n} X_p^+$ is increasing in p for $p \ge n$ and therefore, $M_n := \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]$ exists in $[0,\infty]$. By Fatou's lemma, we know that

$$\mathbb{E}M_n \le \liminf_{p\to\infty} \mathbb{E}\left[\mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right]\right] \le \liminf_{p\to\infty} \mathbb{E}\left[X_p^+\right] = C < \infty$$

which shows M is integrable. By cMCT and the tower property of conditional expectation,

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n} M_{n+1} &= \mathbb{E}_{\mathcal{B}_n} \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] = \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_{n+1}}\left[X_p^+\right] \\ &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] = M_n \text{ a.s.,}\end{aligned}$$

which shows $M = \{M_n\}$ is a martingale.

We now define $Y_n := M_n - X_n$. Using the submartingale property of X^+ implies,

$$\begin{aligned}Y_n = M_n - X_n &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n = \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+\right] - X_n^+ + X_n^- \\ &= \lim_{p\to\infty} \mathbb{E}_{\mathcal{B}_n}\left[X_p^+ - X_n^+\right] + X_n^- \ge 0 \text{ a.s..}\end{aligned}$$

Moreover,

$$\mathbb{E}[Y_{n+1}|\mathcal{B}_n] = \mathbb{E}[M_{n+1} - X_{n+1}|\mathcal{B}_n] = M_n - \mathbb{E}[X_{n+1}|\mathcal{B}_n] \ge M_n - X_n = Y_n$$

wherein we have use M is a martingale in the second equality and X is submartingale the last inequality. ∎

19.3 Stopping Times

Definition 19.20. *Again let* $\{\mathcal{B}_n\}_{n=0}^{\infty}$ *be a filtration on* $(\Omega, \mathcal{B})$ *and assume that* $\mathcal{B} = \mathcal{B}_\infty := \vee_{n=0}^{\infty}\mathcal{B}_n := \sigma\left(\cup_{n=0}^{\infty}\mathcal{B}_n\right)$. *A function,* $\tau : \Omega \to \bar{\mathbb{N}} := \mathbb{N} \cup \{0, \infty\}$ *is said to be a stopping time if* $\{\tau \le n\} \in \mathcal{B}_n$ *for all* $n \in \bar{\mathbb{N}}$. *Equivalently put,* $\tau : \Omega \to \bar{\mathbb{N}}$ *is a stopping time iff the process,* $n \to 1_{\tau \le n}$ *is adapted.*

Lemma 19.21. *Let* $\{\mathcal{B}_n\}_{n=0}^{\infty}$ *be a filtration on* $(\Omega, \mathcal{B})$ *and* $\tau : \Omega \to \bar{\mathbb{N}}$ *be a function. Then the following are equivalent;*

1. τ *is a stopping time.*
2. $\{\tau \le n\} \in \mathcal{B}_n$ *for all* $n \in \mathbb{N}_0$.
3. $\{\tau > n\} = \{\tau \ge n+1\} \in \mathcal{B}_n$ *for all* $n \in \mathbb{N}_0$.
4. $\{\tau = n\} \in \mathcal{B}_n$ *for all* $n \in \mathbb{N}_0$.

Moreover if any of these conditions hold for $n \in \mathbb{N}_0$ *then they also hold for* $n = \infty$.

Proof. (1. $\Longleftrightarrow$ 2.) Observe that if $\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$, then $\{\tau < \infty\} = \cup_{n=1}^{\infty}\{\tau \le n\} \in \mathcal{B}_\infty$ and therefore $\{\tau = \infty\} = \{\tau < \infty\}^c \in \mathcal{B}_\infty$ and hence $\{\tau \le \infty\} = \{\tau < \infty\} \cup \{\tau = \infty\} \in \mathcal{B}_\infty$. Hence in order to check that τ is a stopping time, it suffices to show $\{\tau \le n\} \in \mathcal{B}_n$ for all $n \in \mathbb{N}_0$.

The equivalence of 2., 3., and 4. follows from the identities

$$\{\tau > n\}^c = \{\tau \le n\},$$
$$\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}, \text{ and}$$
$$\{\tau \le n\} = \cup_{k=0}^{n}\{\tau = k\}$$

from which we conclude that 2. $\Longrightarrow$ 3. $\Longrightarrow$ 4. $\Longrightarrow$ 1. ∎

Clearly any constant function, $\tau : \Omega \to \bar{\mathbb{N}}$, is a stopping time. The reader should also observe that if $\mathcal{B}_n = \sigma(X_0, \dots, X_n)$, then $\tau : \Omega \to \bar{\mathbb{N}}$ is a stopping time iff, for each $n \in \mathbb{N}_0$ there exists a measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ such that $1_{\{\tau = n\}} = f_n(X_0, \dots, X_n)$. Here is another common example of a stopping time.

Example 19.22 (First hitting times). Suppose that $X := \{X_n\}_{n=0}^{\infty}$ is an adapted process on the filtered space, $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty})$ and $A \in \mathcal{B}_{\mathbb{R}}$. Then the **first hitting time of** A,

$$\tau := \inf\{n \in \mathbb{N}_0 : X_n \in A\},$$

(with convention that $\inf \emptyset = \infty$) is a stopping time. To see this, observe that

$$\{\tau = n\} = \{X_0 \in A^c, \dots, X_{n-1} \in A^c, X_n \in A\} \in \sigma(X_0, \dots, X_n) \subset \mathcal{B}_n.$$

More generally if σ is a stopping time, then the **first hitting time after** σ,

$$\tau := \inf\{k \ge \sigma : X_k \in A\},$$

is also a stopping time. Indeed,

$$\begin{aligned}\{\tau = n\} &= \{\sigma \le n\} \cap \{X_\sigma \notin A, \dots, X_{n-1} \notin A, X_n \in A\} \\ &= \cup_{0 \le k \le n} \{\sigma = k\} \cap \{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\}\end{aligned}$$

which is in $\mathcal{B}_n$ for all n. Here we use the convention that

$$\{X_k \notin A, \dots, X_{n-1} \notin A, X_n \in A\} = \{X_n \in A\} \text{ if } k = n.$$

On the other hand the last hitting time, $\tau = \sup \{n \in \mathbb{N}_0 : X_n \in A\}$, of a set A is typically not a stopping time. Indeed, in this case

$$\{\tau = n\} = \{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \dots\} \in \sigma(X_n, X_{n+1}, \dots)$$

which typically will not be in $\mathcal{B}_n$.

Proposition 19.23 (New Stopping Times from Old). *Let* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ *be a filtered measure space and suppose* σ, τ, *and* $\{\tau_n\}_{n=1}^\infty$ *are all stopping times. Then*

1. $\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ *are all stopping times.*
2. *If* $\tau_k \uparrow \tau_\infty$ *or* $\tau_k \downarrow \tau_\infty$, *then* τ_∞ *is a stopping time.*
3. *In general,* $\sup_k \tau_k = \lim_{k\to\infty} \max\{\tau_1, \dots, \tau_k\}$ *and* $\inf_k \tau_k = \lim_{k\to\infty} \min\{\tau_1, \dots, \tau_k\}$ *are also stopping times.*

Proof.

1. Since $\{\tau \wedge \sigma > n\} = \{\tau > n\} \cap \{\sigma > n\} \in \mathcal{B}_n$, $\{\tau \vee \sigma \le n\} = \{\tau \le n\} \cap \{\sigma \le n\} \in \mathcal{B}_n$ for all n, and
$$\{\tau + \sigma = n\} = \cup_{k=0}^n \{\tau = k, \sigma = n - k\} \in \mathcal{B}_n$$
for all n, $\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are all stopping times.
2. If $\tau_k \uparrow \tau_\infty$, then $\{\tau_\infty \le n\} = \cap_k \{\tau_k \le n\} \in \mathcal{B}_n$ and so τ_∞ is a stopping time. Similarly, if $\tau_k \downarrow \tau_\infty$, then $\{\tau_\infty > n\} = \cap_k \{\tau_k > n\} \in \mathcal{B}_n$ and so τ_∞ is a stopping time. (Recall that $\{\tau_\infty > n\} = \{\tau_\infty \ge n+1\}$.)
3. This follows from items 1. and 2.

■

Lemma 19.24. *If* τ *is a stopping time, then the processes,* $f_n := 1_{\{\tau \le n\}}$, *and* $f_n := 1_{\{\tau = n\}}$ *are adapted and* $f_n := 1_{\{\tau < n\}}$ *is predictable. Moreover, if* σ *and* τ *are two stopping times, then* $f_n := 1_{\sigma < n \le \tau}$ *is predictable.*

Proof. These are all trivial to prove. For example, if $f_n := 1_{\sigma < n \le \tau}$, then f_n is $\mathcal{B}_{n-1}$ measurable since,

$$\{\sigma < n \le \tau\} = \{\sigma < n\} \cap \{n \le \tau\} = \{\sigma < n\} \cap \{\tau < n\}^c \in \mathcal{B}_{n-1}.$$

■

Notation 19.25 (Stochastic intervals) *If* $\sigma, \tau : \Omega \to \bar{\mathbb{N}}$, *let*

$$(\sigma, \tau] := \left\{(\omega, n) \in \Omega \times \bar{\mathbb{N}} : \sigma(\omega) < n \leq \tau(\omega)\right\}$$

and we will write $1_{(\sigma,\tau]}$ *for the process,* $1_{\sigma<n\leq\tau}$.

Our next goal is to define the "stopped" σ – algebra, $\mathcal{B}_\tau$. To motivate the upcoming definition, suppose $X_n : \Omega \to \mathbb{R}$ are given functions for all $n \in \mathbb{N}_0$, $\mathcal{B}_n := \sigma(X_0, \dots, X_n)$, and $\tau : \Omega \to \mathbb{N}_0$ is a $\mathcal{B}.$ – stopping time. Recalling that a function $Y : \Omega \to \mathbb{R}$ is $\mathcal{B}_n$ measurable iff $Y(\omega) = f_n(X_0(\omega), \dots X_n(\omega))$ for some measurable function, $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$, it is reasonable to suggest that Y is $\mathcal{B}_\tau$ measurable iff $Y(\omega) = f_{\tau(\omega)}\left(X_0(\omega), \dots X_{\tau(\omega)}(\omega)\right)$, where $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ are measurable random variables. If this is the case, then we would have $1_{\tau=n}Y = f_n(X_0, \dots, X_n)$ is $\mathcal{B}_n$ – measurable for all n. Hence we should define $A \subset \Omega$ to be in $\mathcal{B}_\tau$ iff 1_A is $\mathcal{B}_\tau$ measurable iff $1_{\tau=n}1_A$ is $\mathcal{B}_n$ measurable for all n which happens iff $\{\tau = n\} \cap A \in \mathcal{B}_n$ for all n.

Definition 19.26 (Stopped σ – algebra). *Given a stopping time* τ *on a filtered measure space* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty)$ *with* $\mathcal{B}_\infty := \vee_{n=0}^\infty \mathcal{B}_n := \sigma(\cup_{n=0}^\infty \mathcal{B}_n)$, *let*

$$\mathcal{B}_\tau := \{A \subset \Omega : \{\tau = n\} \cap A \in \mathcal{B}_n \text{ for all } n \leq \infty\}. \tag{19.13}$$

Lemma 19.27. *Suppose* σ *and* τ *are stopping times.*

1. *A set,* $A \subset \Omega$ *is in* $\mathcal{B}_\tau$ *iff* $A \cap \{\tau \leq n\} \in \mathcal{B}_n$ *for all* $n \leq \infty$.
2. $\mathcal{B}_\tau$ *is a sub-*σ*-algebra of* $\mathcal{B}_\infty$.
3. *If* $\sigma \leq \tau$, *then* $\mathcal{B}_\sigma \subset \mathcal{B}_\tau$.

Proof. 1. Since

$$A \cap \{\tau \leq n\} = \cup_{k\leq n}[A \cap \{\tau \leq k\}] \text{ and}$$
$$A \cap \{\tau = n\} = [A \cap \{\tau \leq n\}] \setminus [A \cap \{\tau \leq n-1\}],$$

it easily follows that $A \subset \Omega$ is in $\mathcal{B}_\tau$ iff $A \cap \{\tau \leq n\} \in \mathcal{B}_n$ for all $n \leq \infty$.

2. Since $\Omega \cap \{\tau \leq n\} = \{\tau \leq n\} \in \mathcal{B}_n$ for all n, it follows that $\Omega \in \mathcal{B}_\tau$. If $A \in \mathcal{B}_\tau$, then, for all $n \in \mathbb{N}_0$,

$$A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus A = \{\tau \leq n\} \setminus [A \cap \{\tau \leq n\}] \in \mathcal{B}_n.$$

This shows $A^c \in \mathcal{B}_\tau$. Similarly if $\{A_k\}_{k=1}^\infty \subset \mathcal{B}_\tau$, then

$$\{\tau \leq n\} \cap (\cap_{k=1}^\infty A_k) = \cap_{k=1}^\infty (\{\tau \leq n\} \cap A_k) \in \mathcal{B}_n$$

and hence $\cap_{k=1}^\infty A_k \in \mathcal{B}_\tau$. This completes the proof the $\mathcal{B}_\tau$ is a σ – algebra. Since $A = A \cap \{\tau \leq \infty\}$, it also follows that $\mathcal{B}_\tau \subset \mathcal{B}_\infty$.

3. Now suppose that $\sigma \leq \tau$ and $A \in \mathcal{B}_\sigma$. Since $A \cap \{\sigma \leq n\}$ and $\{\tau \leq n\}$ are in $\mathcal{B}_n$ for all $n \leq \infty$, we find

$$A \cap \{\tau \leq n\} = [A \cap \{\sigma \leq n\}] \cap \{\tau \leq n\} \in \mathcal{B}_n \ \forall\ n \leq \infty$$

which shows $A \in \mathcal{B}_\tau$. ■

Proposition 19.28 ($\mathcal{B}_\tau$ – measurable random variables). *Let $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty})$ be a filtered measure space. Let τ be a stopping time and $Z : \Omega \to \mathbb{R}$ be a function. Then the following are equivalent;*

1. *Z is $\mathcal{B}_\tau$ – measurable,*
2. *$1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$,*
3. *$1_{\{\tau = n\}} Z$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$.*
4. *There exists, $Y_n : \Omega \to \mathbb{R}$ which are $\mathcal{B}_n$ – measurable for all $n \le \infty$ such that*
$$Z = Y_\tau = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau = n\}} Y_n.$$

Proof. 1. $\implies$ 2. By definition, if $A \in \mathcal{B}_\tau$, then $1_{\{\tau \le n\}} 1_A = 1_{\{\tau \le n\} \cap A}$ is $\mathcal{B}_n$ – measurable for all $n \le \infty$. Consequently any simple $\mathcal{B}_\tau$ – measurable function, Z, satisfies $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all n. So by the usual limiting argument (Theorem 6.32), it follows that $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable for all n for any $\mathcal{B}_\tau$ – measurable function, Z.

2. $\implies$ 3. This property follows from the identity,
$$1_{\{\tau = n\}} Z = 1_{\{\tau \le n\}} Z - 1_{\{\tau < n\}} Z.$$

3. $\implies$ 4. Simply take $Y_n = 1_{\{\tau = n\}} Z$.

4. $\implies$ 1. Since $Z = \sum_{n \in \bar{\mathbb{N}}} 1_{\{\tau = n\}} Y_n$, it suffices to show $1_{\{\tau = n\}} Y_n$ is $\mathcal{B}_\tau$ – measurable if Y_n is $\mathcal{B}_n$ – measurable. Further, by the usual limiting arguments using Theorem 6.32, it suffices to assume that $Y_n = 1_A$ for some $A \in \mathcal{B}_n$. In this case $1_{\{\tau = n\}} Y_n = 1_{A \cap \{\tau = n\}}$. Hence we must show $A \cap \{\tau = n\} \in \mathcal{B}_\tau$ which indeed is true because
$$A \cap \{\tau = n\} \cap \{\tau = k\} = \begin{cases} \emptyset \in \mathcal{B}_k & \text{if } k \ne n \\ A \cap \{\tau = n\} \in \mathcal{B}_k & \text{if } k = n \end{cases}.$$

Alternatively proof for 1. $\implies$ 2. If Z is $\mathcal{B}_\tau$ measurable, then $\{Z \in B\} \cap \{\tau \le n\} \in \mathcal{B}_n$ for all $n \le \infty$ and $B \in \mathcal{B}_{\mathbb{R}}$. Hence if $B \in \mathcal{B}_{\mathbb{R}}$ with $0 \notin B$, then
$$\left\{1_{\{\tau \le n\}} Z \in B\right\} = \{Z \in B\} \cap \{\tau \le n\} \in \mathcal{B}_n \text{ for all } n$$
and similarly,
$$\left\{1_{\{\tau \le n\}} Z = 0\right\}^c = \left\{1_{\{\tau \le n\}} Z \ne 0\right\} = \{Z \ne 0\} \cap \{\tau \le n\} \in \mathcal{B}_n \text{ for all } n.$$
From these two observations, it follows that $\left\{1_{\{\tau \le n\}} Z \in B\right\} \in \mathcal{B}_n$ for all $B \in \mathcal{B}_{\mathbb{R}}$ and therefore, $1_{\{\tau \le n\}} Z$ is $\mathcal{B}_n$ – measurable. ■

Lemma 19.29 ($\mathcal{B}_\sigma$ – conditioning). *Suppose σ is a stopping time and $Z \in L^1(\Omega, \mathcal{B}, P)$ or $Z \ge 0$, then*
$$\mathbb{E}[Z|\mathcal{B}_\sigma] = \sum_{n \le \infty} 1_{\sigma = n} \mathbb{E}[Z|\mathcal{B}_n] = Y_\sigma \tag{19.14}$$

where

$$Y_n := \mathbb{E}[Z|\mathcal{B}_n] \text{ for all } n \in \bar{\mathbb{N}}. \tag{19.15}$$

Proof. By Proposition 19.28, Y_σ is $\mathcal{B}_\sigma$ – measurable. Moreover if Z is integrable, then

$$\begin{aligned}\sum_{n\le\infty} \mathbb{E}\left[1_{\{\sigma=n\}}|Y_n|\right] &= \sum_{n\le\infty} \mathbb{E}1_{\{\sigma=n\}}|\mathbb{E}[Z|\mathcal{B}_n]| \\ &\le \sum_{n\le\infty} \mathbb{E}\left[1_{\{\sigma=n\}}\mathbb{E}[|Z||\mathcal{B}_n]\right] \\ &= \sum_{n\le\infty} \mathbb{E}\left[\mathbb{E}\left[1_{\{\sigma=n\}}|Z||\mathcal{B}_n\right]\right] \\ &= \sum_{n\le\infty} \mathbb{E}\left[1_{\{\sigma=n\}}|Z|\right] = \mathbb{E}|Z| < \infty \end{aligned} \tag{19.16}$$

and therefore

$$\begin{aligned}\mathbb{E}|Y_\sigma| &= \mathbb{E}\left|\sum_{n\le\infty}\left[1_{\{\sigma=n\}}Y_n\right]\right| \\ &\le \sum_{n\le\infty} \mathbb{E}\left[1_{\{\sigma=n\}}|Y_n|\right] \le \mathbb{E}|Z| < \infty.\end{aligned}$$

Furthermore if $A \in \mathcal{B}_\sigma$, then

$$\begin{aligned}\mathbb{E}[Z:A] &= \sum_{n\le\infty} \mathbb{E}[Z:A\cap\{\sigma=n\}] = \sum_{n\le\infty} \mathbb{E}[Y_n:A\cap\{\sigma=n\}] \\ &= \sum_{n\le\infty} \mathbb{E}\left[1_{\{\sigma=n\}}Y_n:A\right] = \mathbb{E}\left[\sum_{n\le\infty} 1_{\{\sigma=n\}}Y_n:A\right] \\ &= \mathbb{E}[Y_\sigma:A],\end{aligned}$$

wherein the interchange of the sum and the expectation in the second to last equality is justified by the estimate in 19.16 or by the fact that everything in sight is positive when $Z \ge 0$. ■

Exercise 19.2. Suppose σ and τ are two stopping times. Show;

1. $\{\sigma < \tau\}$, $\{\sigma = \tau\}$, and $\{\sigma \le \tau\}$ are all in $\mathcal{B}_\sigma \cap \mathcal{B}_\tau$,
2. $\mathcal{B}_{\sigma\wedge\tau} = \mathcal{B}_\sigma \cap \mathcal{B}_\tau$,
3. $\mathcal{B}_{\sigma\vee\tau} = \mathcal{B}_\sigma \vee \mathcal{B}_\tau := \sigma(\mathcal{B}_\sigma \cup \mathcal{B}_\tau)$, and
4. $(\mathcal{B}_\sigma)_{\{\sigma\le\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$ and $(\mathcal{B}_\sigma)_{\{\sigma<\tau\}} \subset \mathcal{B}_{\sigma\wedge\tau}$.

Recall that

$$(\mathcal{B}_\sigma)_{\{\sigma\le\tau\}} = \{A \cap \{\sigma \le \tau\} : A \in \mathcal{B}_\sigma\}.$$

Exercise 19.3 (Tower Property II). Let $X \in L^1(\Omega, \mathcal{B}, P)$ or $X : \Omega \to [0,\infty]$ be a $\mathcal{B}$ – measurable function. Then given **any** two stopping times, σ and τ, show

$$\mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau}X = \mathbb{E}_{\mathcal{B}_\tau}\mathbb{E}_{\mathcal{B}_\sigma}X = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}}X. \tag{19.17}$$

(**Hints:** 1. It suffices to consider the case where $X \geq 0$. 2. Make use of Exercise 19.2, Lemma 19.29 and the basic properties of conditional expectations. If you want to be sophisticated you may also want to use the localization Lemma 18.19 – but it can be avoided if you choose.)

Exercise 19.4. Show, by example, that it is not necessarily true that

$$\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2} = \mathbb{E}_{\mathcal{G}_1\wedge\mathcal{G}_2}$$

for arbitrary $\mathcal{G}_1$ and $\mathcal{G}_2$ – sub-sigma algebras of $\mathcal{B}$.

Hint: it suffices to take $(\Omega, \mathcal{B}, P)$ with $\Omega = \{1,2,3\}$, $\mathcal{B} = 2^\Omega$, and $P(\{j\}) = \frac{1}{3}$ for $j = 1,2,3$.

19.4 Stochastic Integrals and Optional Stopping

Notation 19.30 *Suppose that $\{u_n\}_{n=1}^\infty$ and $\{x_n\}_{n=0}^\infty$ are two sequences of numbers, let udx denote the sequence of numbers defined by*

$$(u\cdot\Delta x)_n = \sum_{j=1}^n u_j(x_j - x_{j-1}) = \sum_{j=1}^n u_j\Delta_j x \text{ for } n \geq 1.$$

For a gambling interpretation of $(u\cdot\Delta x)_n$, let x_j represent the price of a stock at time j. Suppose that you, the investor, then buys u_{j-1} shares at time $j-1$ and then sells these shares back at time j. With this interpretation, $u_{j-1}\Delta_j x$ represents your profit (or loss if negative) in the time interval form $j-1$ to j and $(u\cdot\Delta x)_n$ represents your profit (or loss) from time 0 to time n. By the way, if you want to buy 5 shares of the stock at time $n=0$ and then sell them all at time n, you would take $u_k = 5\cdot 1_{k<n}$.

Example 19.31. Suppose that $0 \leq \sigma \leq \tau$ where $\sigma,\tau \in \bar{\mathbb{N}}_0$ and let $u_n := 1_{\sigma<n\leq\tau}$. Then

$$\begin{aligned}(u\cdot\Delta x)_n &= \sum_{j=1}^n 1_{\sigma<j\leq\tau}(x_j - x_{j-1}) = \sum_{j=1}^\infty 1_{\sigma<j\leq\tau\wedge n}(x_j - x_{j-1})\\ &= x_{\tau\wedge n} - x_{\sigma\wedge n}.\end{aligned}$$

Proposition 19.32 (The Discrete Stochastic Integral). *Let $X = \{X_n\}_{n=0}^\infty$ be an adapted integrable process, i.e. $\mathbb{E}|X_n| < \infty$ for all n. If X is a martingale and $\{U_n\}_{n=1}^\infty$ is a predictable sequence of bounded random variables, then $\{(U\cdot\Delta X)_n\}_{n=1}^\infty$ is still a martingale. If $X := \{X_n\}_{n=0}^\infty$ is a* *submartingale (supermartingale) (necessarily real valued) and* $U_n \geq 0$, *then* $\{(U \cdot \Delta X)_n\}_{n=1}^{\infty}$ *is a submartingale (supermartingale).*

Conversely if X *is an adapted process of integrable functions such that* $\mathbb{E}\left[(U \cdot \Delta X)_n\right] = 0$ *for all bounded predictable processes,* $\{U_n\}_{n=1}^{\infty}$, *then* X *is a martingale. Similarly if* X *is real valued adapted process such that*

$$\mathbb{E}\left[(U \cdot \Delta X)_n\right] \overset{\leq}{\underset{\geq}{=}} 0 \tag{19.18}$$

for all n *and for all bounded, non-negative predictable processes,* U, *then* X *is a supermartingale, martingale, or submartingale respectively.*

Proof. For any adapted process X, we have

$$\begin{aligned}\mathbb{E}\left[(U \cdot \Delta X)_{n+1} | \mathcal{B}_n\right] &= \mathbb{E}\left[(U \cdot \Delta X)_n + U_{n+1}\left(X_{n+1} - X_n\right) | \mathcal{B}_n\right] \\ &= (U \cdot \Delta X)_n + U_{n+1}\mathbb{E}\left[\left(X_{n+1} - X_n\right) | \mathcal{B}_n\right].\end{aligned} \tag{19.19}$$

The first assertions easily follow from this identity.

Now suppose that X is an adapted process of integrable functions such that $\mathbb{E}\left[(U \cdot \Delta X)_n\right] = 0$ for all bounded predictable processes, $\{U_n\}_{n=1}^{\infty}$. Taking expectations of Eq. (19.19) then allows us to conclude that

$$\mathbb{E}\left[U_{n+1}\mathbb{E}\left[\left(X_{n+1} - X_n\right) | \mathcal{B}_n\right]\right] = 0$$

for all bounded $\mathcal{B}_n$ – measurable random variables, U_{n+1}. Taking $U_{n+1} := \operatorname{sgn}(\mathbb{E}\left[\left(X_{n+1} - X_n\right) | \mathcal{B}_n\right])$ shows $\left|\mathbb{E}\left[\left(X_{n+1} - X_n\right) | \mathcal{B}_n\right]\right| = 0$ a.s. and hence X is a martingale. Similarly, if for all non-negative, predictable U, Eq. (19.18) holds for all $n \geq 1$, and $U_n \geq 0$, then taking $A \in \mathcal{B}_n$ and $U_k = \delta_{k,n+1} 1_A$ in Eq. (19.12) allows us to conclude that

$$\mathbb{E}\left[X_{n+1} - X_n : A\right] = \mathbb{E}\left[(U \cdot \Delta X)_{n+1}\right] \overset{\leq}{\underset{\geq}{=}} 0,$$

i.e. X is a supermartingale, martingale, or submartingale respectively. ∎

Example 19.33. Suppose that $\{X_n\}_{n=0}^{\infty}$ are mean zero independent integrable random variables and $f_n : \mathbb{R}^n \to \mathbb{R}$ are bounded measurable functions. Then

$$Y_n := \sum_{j=1}^{n} f_n\left(X_0, \ldots, X_{n-1}\right)\left(X_n - X_{n-1}\right) \tag{19.20}$$

defines a martingale sequence.

Notation 19.34 *Given an adapted process,* X, *and a stopping time* τ, *let* $X_n^{\tau} := X_{\tau \wedge n}$. *We call* $X^{\tau} := \{X_n^{\tau}\}_{n=0}^{\infty}$ *the* ***process*** X ***stopped by*** τ.

Theorem 19.35 (Optional stopping theorem). *Suppose* $X = \{X_n\}_{n=0}^{\infty}$ *is a supermartingale, martingale, or submartingale and* τ *is a stopping time, then* X^{τ} *is a* $\{\mathcal{B}_n\}_{n=0}^{\infty}$ *– supermartingale, martingale, or submartingale respectively. This valid if either* $\mathbb{E}|X_n| < \infty$ *for all* n *or if* $X_n \geq 0$ *for all* n.

Proof. First proof. Since $1_{\tau\le n}X_\tau = \sum_{k=0}^n 1_{\tau=n}X_n$ is $\mathcal{B}_n$ measurable, $\{\tau > n\} \in \mathcal{B}_n$, and

$$X_{\tau\wedge(n+1)} = 1_{\tau\le n}X_\tau + 1_{\tau>n}X_{n+1},$$

we have

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_n}\left[X_{\tau\wedge(n+1)}\right] &= 1_{\tau\le n}X_\tau + 1_{\tau>n}\mathbb{E}_{\mathcal{B}_n}X_{n+1} \\ &\overset{\le}{\underset{\ge}{=}} 1_{\tau\le n}X_\tau + 1_{\tau>n}X_n = X_{\tau\wedge n}.\end{aligned}$$

Second proof in case $\mathbb{E}|X_n| < \infty$. Let $U_k := 1_{0<k\le\tau}$ for $k = 1, 2, \ldots$. Then U is a bounded predictable process and

$$(U\cdot\Delta X)_n = \sum_{k\le n} 1_{0<k\le\tau}\Delta_k X = \sum_{0<k\le\tau\wedge n}\Delta_k X = X_{\tau\wedge n} - X_0.$$

Therefore, by Proposition 19.32, $X_n^\tau = X_0 + (U\cdot\Delta X)_n$ is (respectively) a supermartingale, martingale, or submartingale.

Third proof. See Remark 19.37 below. ■

Theorem 19.36 (Optional sampling theorem I). *Suppose that σ and τ are two stopping times and τ is bounded, i.e. there exists $N \in \mathbb{N}$ such that $\tau \le N < \infty$ a.s. If $X = \{X_n\}_{n=0}^\infty$ is an integrable supermartingale, martingale, or submartingale, then X_τ is integrable and*

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] \overset{\le}{\underset{\ge}{=}} X_{\sigma\wedge\tau} \text{ a.s.} \tag{19.21}$$

respectively[2] *from top to bottom. Moreover, Eq. (19.21) is valid with no integrability assumptions on X provided $X_n \ge 0$ a.s. for all $n < \infty$.*

Proof. Since

$$|X_\tau| = \left|\sum_{0\le k\le\tau} 1_{\tau=k}X_k\right| \le \sum_{0\le k\le\tau} 1_{\tau=k}|X_k| \le \sum_{0\le k\le N}|X_k|,$$

if $X_n \in L^1(\Omega,\mathcal{B},P)$ for all n we see that $\mathbb{E}|X_\tau| \le \sum_{0\le k\le N}\mathbb{E}|X_k| < \infty$. Hence it remains to prove Eq. (19.21) in case $X_n \ge 0$ or $X_n \in L^1(\Omega,\mathcal{B},P)$ for all n.

According to Lemma 19.29

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] = \sum_{n=0}^\infty 1_{\sigma=n}\mathbb{E}\left[X_\tau|\mathcal{B}_n\right]. \tag{19.22}$$

[2] This is the natural generalization of Eq. (19.3) to the stopping time setting.

On the other hand we know X^τ is a supermartingale, martingale, or submartingale respectively and therefore, for any $n < \infty$ and $m \geq \max(n, N)$ we have

$$\mathbb{E}\left[X_\tau|\mathcal{B}_n\right] = \mathbb{E}\left[X_m^\tau|\mathcal{B}_n\right] \overset{\leq}{\underset{\geq}{=}} X_n^\tau = X_{\tau\wedge n}.$$

Combining this equation with Eq. (19.22) shows

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] \overset{\leq}{\underset{\geq}{=}} \sum_{n=0}^{\infty} 1_{\sigma=n} X_{\tau\wedge n} = X_{\tau\wedge\sigma}.$$

This completes the proof. Nevertheless we will give two more proofs of Eq. (19.22) under the assumption that $X_n \in L^1(\Omega, \mathcal{B}, P)$ for all n.

First alternative proof. First suppose X is a martingale in which case $X_n = \mathbb{E}_{\mathcal{B}_n} X_N$ for all $n \leq N$ and hence

$$X_\tau = \sum_{n\leq N} 1_{\tau=n} X_n = \sum_{n\leq N} 1_{\tau=n}\mathbb{E}_{\mathcal{B}_n} X_N = \sum_{n\leq\infty} 1_{\tau=n}\mathbb{E}_{\mathcal{B}_n} X_N = \mathbb{E}_{\mathcal{B}_\tau} X_N.$$

Therefore, by Exercise 19.3

$$\mathbb{E}_{\mathcal{B}_\sigma} X_\tau = \mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau} X_N = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}} X_N = X_{\sigma\wedge\tau}.$$

Now suppose that X is a submartingale. By the Doob decomposition (Lemma 19.15), $X_n = M_n + A_n$ where M is a martingale and A is an increasing predictable process. In this case we have

$$\begin{aligned}\mathbb{E}_{\mathcal{B}_\sigma} X_\tau &= \mathbb{E}_{\mathcal{B}_\sigma} M_\tau + \mathbb{E}_{\mathcal{B}_\sigma} A_\tau = M_{\sigma\wedge\tau} + \mathbb{E}_{\mathcal{B}_\sigma} A_\tau \\ &\geq M_{\sigma\wedge\tau} + \mathbb{E}_{\mathcal{B}_\sigma} A_{\sigma\wedge\tau} = M_{\sigma\wedge\tau} + A_{\sigma\wedge\tau} = X_{\sigma\wedge\tau}.\end{aligned}$$

The supermartingale case follows from the submartingale result just proved applied to $-X$.

Second alternative proof. Let $A \in \mathcal{B}_\sigma$ and $U_n := 1_A \cdot 1_{\sigma<n\leq\tau}$. Then U is predictable since

$$A \cap \{\sigma < n \leq \tau\} = (A \cap \{\sigma < n\}) \cap \{n \leq \tau\} \in \mathcal{B}_{n-1} \text{ for all } n.$$

Let us also observe that

$$\begin{aligned}(U \cdot \Delta X)_n &= \sum_{k\leq n} 1_A \cdot 1_{\sigma<k\leq\tau}\Delta_k X = \sum 1_A \cdot 1_{\sigma\wedge\tau<k\leq\tau\wedge n}\Delta_k X \\ &= 1_A\left(X_{\tau\wedge n} - X_{\sigma\wedge\tau}\right) \text{ for all } n \geq 1.\end{aligned}$$

By Proposition 19.32, $(U \cdot \Delta X)$ is a supermartingale, martingale, or submartingale respectively and hence

$$\mathbb{E}\left[1_A\left(X_\tau - X_{\sigma\wedge\tau}\right)\right] = \mathbb{E}\left[1_A\left(X_{\tau\wedge N} - X_{\sigma\wedge\tau}\right)\right] = \mathbb{E}\left[(U \cdot \Delta X)_N\right] \overset{\leq}{\underset{\geq}{=}} 0 \text{ respectively.}$$

Since $A \in \mathcal{B}_\sigma$ is arbitrary and $X_{\sigma\wedge\tau}$ is $\mathcal{B}_\sigma$ – measurable (in fact $\mathcal{B}_{\sigma\wedge\tau}$ – measurable), Eq. (19.21) has been proved. ∎

Remark 19.37. Theorem 19.36 can be used to give a simple proof of the Optional stopping Theorem 19.35. For example, if $X = \{X_n\}_{n=0}^{\infty}$ is a submartingale and τ is a stopping time, then

$$\mathbb{E}_{\mathcal{B}_n} X_{\tau\wedge(n+1)} \geq X_{[\tau\wedge(n+1)]\wedge n} = X_{\tau\wedge n},$$

i.e. X^{τ} is a submartingale.

19.5 Submartingale Inequalities

For a process, $X = \{X_n\}_{n=0}^{\infty}$ let

$$X_N^* := \max\{|X_0|, \ldots, |X_N|\}. \tag{19.23}$$

19.5.1 Maximal Inequalities

Proposition 19.38 (Maximal Inequalities of Bernstein and Lévy). *Let* $\{X_n\}$ *be a submartingale on a filtered probability space,* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$. *Then*[3] *for any* $a \geq 0$ *and* $N \in \mathbb{N}$,

$$aP\left(\max_{n\leq N} X_n \geq a\right) \leq \mathbb{E}\left[X_N : \max_{n\leq N} X_n \geq a\right] \leq \mathbb{E}\left[X_N^+\right], \tag{19.24}$$

$$aP\left(\min_{n\leq N} X_n \leq -a\right) \leq \mathbb{E}\left[X_N : \min_{k\leq N} X_k > -a\right] - \mathbb{E}[X_0] \tag{19.25}$$

$$\leq \mathbb{E}\left[X_N^+\right] - \mathbb{E}[X_0], \tag{19.26}$$

and

$$aP(X_N^* \geq a) \leq 2\mathbb{E}\left[X_N^+\right] - \mathbb{E}[X_0]. \tag{19.27}$$

Proof. Initially let X be **any** integrable adapted process and τ be the stopping time defined by, $\tau := \inf\{n : X_n \geq a\}$. Since $X_\tau \geq a$ on

$$\{\tau \leq N\} = \left\{\max_{n\leq N} X_n \geq a\right\}, \tag{19.28}$$

we have

$$aP\left(\max_{n\leq N} X_n \geq a\right) = \mathbb{E}[a : \tau \leq N] \leq \mathbb{E}[X_\tau : \tau \leq N] \tag{19.29}$$

$$= \mathbb{E}[X_N : \tau \leq N] - \mathbb{E}[X_N - X_\tau : \tau \leq N]$$

$$= \mathbb{E}[X_N : \tau \leq N] - \mathbb{E}[X_N - X_{\tau\wedge N}]. \tag{19.30}$$

[3] The first inequality is the most important.

Let me emphasize again that in deriving Eq. (19.30), we have **not** used any special properties (not even adaptedness) of X.

If X is now assumed to be a submartingale, by the optional sampling Theorem 19.36, $\mathbb{E}_{\mathcal{B}_{\tau\wedge N}} X_N \geq X_{\tau\wedge N}$ and in particular $\mathbb{E}[X_N - X_{\tau\wedge N}] \geq 0$. Combining this observation with Eq. (19.30) and Eq. (19.28) gives Eq. (19.24). (Alternatively, since $\{\tau \leq N\} \in B_{\tau\wedge N}$, it follows by optional sampling that

$$\mathbb{E}[X_\tau : \tau \leq N] = \mathbb{E}[X_{\tau\wedge N} : \tau \leq N] \leq \mathbb{E}[X_N : \tau \leq N]$$

which combined with Eq. (19.29) gives Eq. (19.24).)

Secondly we may apply Eq. (19.30) with X_n replaced by $-X_n$ to find

$$\begin{aligned} aP\left(\min_{n\leq N} X_n \leq -a\right) &= aP\left(-\min_{n\leq N} X_n \geq a\right) = aP\left(\max_{n\leq N}(-X_n) \geq a\right) \\ &\leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] \end{aligned} \tag{19.31}$$

where now,

$$\tau := \inf\{n : -X_n \geq a\} = \inf\{n : X_n \leq -a\}.$$

By the optional sampling Theorem 19.36, $\mathbb{E}[X_{\tau\wedge N} - X_0] \geq 0$ and adding this to right side of Eq. (19.31) gives the estimate

$$\begin{aligned} aP\left(\min_{n\leq N} X_n \leq -a\right) &\leq -\mathbb{E}[X_N : \tau \leq N] + \mathbb{E}[X_N - X_{\tau\wedge N}] + \mathbb{E}[X_{\tau\wedge N} - X_0] \\ &\leq \mathbb{E}[X_N - X_0] - \mathbb{E}[X_N : \tau \leq N] \\ &= \mathbb{E}[X_N : \tau > N] - \mathbb{E}[X_0] \\ &= \mathbb{E}\left[X_N : \min_{k\leq N} X_k > -a\right] - \mathbb{E}[X_0] \end{aligned}$$

which proves Eq. (19.25) and hence Eq. (19.26). Adding Eqs. (19.24) and (19.26) gives the estimate in Eq. (19.27). ■

Remark 19.39. It is of course possible to give a direct proof of Proposition 19.38. For example,

$$\begin{aligned} \mathbb{E}\left[X_N : \max_{n\leq N} X_n \geq a\right] &= \sum_{k=1}^N \mathbb{E}[X_N : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[X_k : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &\geq \sum_{k=1}^N \mathbb{E}[a : X_1 < a, \dots, X_{k-1} < a, X_k \geq a] \\ &= aP\left(\max_{n\leq N} X_n \geq a\right) \end{aligned}$$

which proves Eq. (19.24).

Example 19.40. Let $\{X_n\}$ be a sequence of independent random variables with mean zero, $S_n := X_1 + \cdots + X_n$, and $S_n^* = \max_{j \le n} |S_j|$. Since $\{S_n\}_{n=1}^\infty$ is a martingale and $\{|S_n|^p\}_{n=1}^\infty$ is an (possibly extended) submartingale for any $p \in [1, \infty)$. Therefore an application of Eq. (19.24) of Proposition 19.38 show

$$P(S_N^* \ge \alpha) = P\left(S_N^{*p} \ge \alpha^p\right) \le \frac{1}{\alpha^p} \mathbb{E}\left[|S_N|^p : S_N^* \ge \alpha\right].$$

When $p = 2$, this is Kolmogorov's Inequality, see Theorem 12.28.

Lemma 19.41. *Suppose that X and Y are two non-negative random variables such that $P(Y \ge y) \le \frac{1}{y} \mathbb{E}[X : Y \ge y]$ for all $y > 0$. Then for all $p \in (1, \infty)$,*

$$\mathbb{E}Y^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p. \tag{19.32}$$

Proof. We will begin by proving Eq. (19.32) under the additional assumption that $Y \in L^p(\Omega, \mathcal{B}, P)$. Since

$$\begin{aligned}
\mathbb{E}Y^p &= p\mathbb{E} \int_0^\infty 1_{y \le Y} \cdot y^{p-1} dy = p \int_0^\infty \mathbb{E}\left[1_{y \le Y}\right] \cdot y^{p-1} dy \\
&= p \int_0^\infty P(Y \ge y) \cdot y^{p-1} dy \le p \int_0^\infty \frac{1}{y} \mathbb{E}[X : Y \ge y] \cdot y^{p-1} dy \\
&= p\mathbb{E} \int_0^\infty X 1_{y \le Y} \cdot y^{p-2} dy = \frac{p}{p-1} \mathbb{E}\left[XY^{p-1}\right].
\end{aligned}$$

Now apply Hölder's inequality, with $q = p(p-1)^{-1}$, to find

$$\mathbb{E}\left[XY^{p-1}\right] \le \|X\|_p \cdot \left\|Y^{p-1}\right\|_q = \|X\|_p \cdot \left[\mathbb{E}|Y|^p\right]^{1/q}.$$

Combining thew two inequalities shows and solving for $\|Y\|_p$ shows $\|Y\|_p \le \frac{p}{p-1} \|X\|_p$ which proves Eq. (19.32) under the additional restriction of Y being in $L^p(\Omega, \mathcal{B}, P)$.

To remove the integrability restriction on Y, for $M > 0$ let $Z := Y \wedge M$ and observe that

$$P(Z \ge y) = P(Y \ge y) \le \frac{1}{y} \mathbb{E}[X : Y \ge y] = \frac{1}{y} \mathbb{E}[X : Z \ge y] \text{ if } y \le M$$

while

$$P(Z \ge y) = 0 = \frac{1}{y} \mathbb{E}[X : Z \ge y] \text{ if } y > M.$$

Since Z is bounded, the special case just proved shows

$$\mathbb{E}\left[(Y \wedge M)^p\right] = \mathbb{E}Z^p \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X^p.$$

We may now use the MCT to pass to the limit, $M \uparrow \infty$, and hence conclude that Eq. (19.32) holds in general. ■

Corollary 19.42 (Doob's Inequality). *If $X = \{X_n\}_{n=0}^{\infty}$ be a non-negative submartingale and $1 < p < \infty$, then*

$$\mathbb{E}X_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}X_N^p. \tag{19.33}$$

Proof. Equation 19.33 follows by applying Lemma 19.41 with the aid of Proposition 19.38. ∎

Corollary 19.43 (Doob's Inequality). *If $\{M_n\}_{n=0}^{\infty}$ is a martingale and $1 < p < \infty$, then for all $a > 0$,*

$$P\left(M_N^* \ge a\right) \le \frac{1}{a}\mathbb{E}\left[|M|_N : M_N^* \ge a\right] \le \frac{1}{a}\mathbb{E}\left[|M_N|\right] \tag{19.34}$$

and

$$\mathbb{E}M_N^{*p} \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left|M_N\right|^p. \tag{19.35}$$

Proof. By the conditional Jensen's inequality, it follows that $X_n := |M_n|$ is a submartingale. Hence Eq. (19.34) follows from Eq. (19.24) and Eq. (19.35) follows from Eq. (19.33). ∎

19.5.2 Upcrossing Inequalities and Convergence Results

Given a function, $\mathbb{N}_0 \ni n \to X_n \in \mathbb{R}$ and $-\infty < a < b < \infty$, let

$$\begin{aligned} \tau_0 = 0,\ \tau_1 &= \inf\{n \ge \tau_0 : X_n \le a\} \\ \tau_2 = \inf\{n \ge \tau_1 : X_n \ge b\},\ \tau_3 &:= \inf\{n \ge \tau_2 : X_n \le a\} \\ &\vdots \\ \tau_{2k} = \inf\{n \ge \tau_{2k-1} : X_n \ge b\},\ \tau_{2k+1} &:= \inf\{n \ge \tau_{2k} : X_n \le a\} \\ &\vdots \end{aligned} \tag{19.36}$$

with the usual convention that $\inf \emptyset = \infty$ in the definitions above, see Figures 19.1 and 19.2. Observe that $\tau_{n+1} \ge \tau_n + 1$ for all $n \ge 1$ and hence $\tau_n \ge n - 1$ for all $n \ge 1$. Further, for each $N \in \bar{\mathbb{N}}$ let

$$U_N^X(a,b) = \max\{k : \tau_{2k} \le N\}$$

be the **number of upcrossings of X across** $[a,b]$ in the time interval, $[0,N]$.

Lemma 19.44. *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a sequence of extended real numbers such that $U_{\infty}^X(a,b) < \infty$ for all $a, b \in \mathbb{Q}$ with $a < b$. Then $X_\infty := \lim_{n\to\infty} X_n$ exists in $\bar{\mathbb{R}}$.*

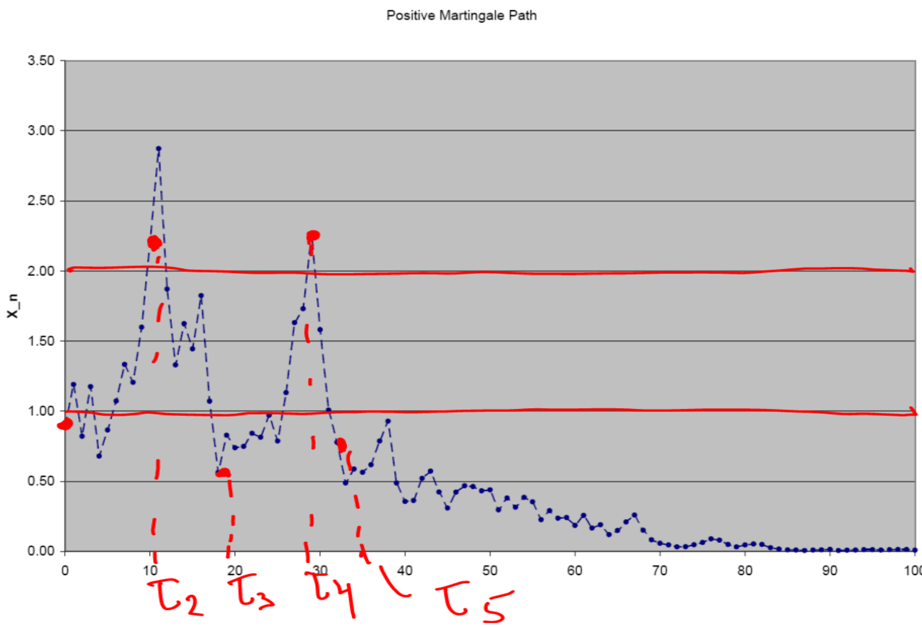


Fig. 19.1. A sample path of a positive martingale with crossing levels, $a = 1$ and $b = 2$ marked off.

Proof. If $\lim_{n\to\infty} X_n$ does not exists in $\bar{\mathbb{R}}$, then there would exists $a, b \in \mathbb{Q}$ such that

$$\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n$$

and for this choice of a and b, we must have $X_n < a$ and $X_n > b$ infinitely often. Therefore, $U_\infty^X(a, b) = \infty$. ■

Theorem 19.45 (Doob's Upcrossing Inequality – buy low sell high). *If $\{X_n\}_{n=0}^\infty$ is a submartingale and $-\infty < a < b < \infty$, then for all $N \in \mathbb{N}$,*

$$\mathbb{E}\left[U_N^X(a, b)\right] \le \frac{1}{b-a}\left[\mathbb{E}(X_N - a)_+ - \mathbb{E}(X_0 - a)_+\right].$$

Proof. We first suppose that $X_n \ge 0$, $a = 0$ and $b > 0$. Let

$$\tau_0 = 0, \ \tau_1 = \inf\{n \ge \tau_0 : X_n = 0\}$$
$$\tau_2 = \inf\{n \ge \tau_1 : X_n \ge b\}, \ \tau_3 := \inf\{n \ge \tau_2 : X_n = 0\}$$
$$\vdots$$
$$\tau_{2k} = \inf\{n \ge \tau_{2k-1} : X_n \ge b\}, \ \tau_{2k+1} := \inf\{n \ge \tau_{2k} : X_n = 0\}$$
$$\vdots$$

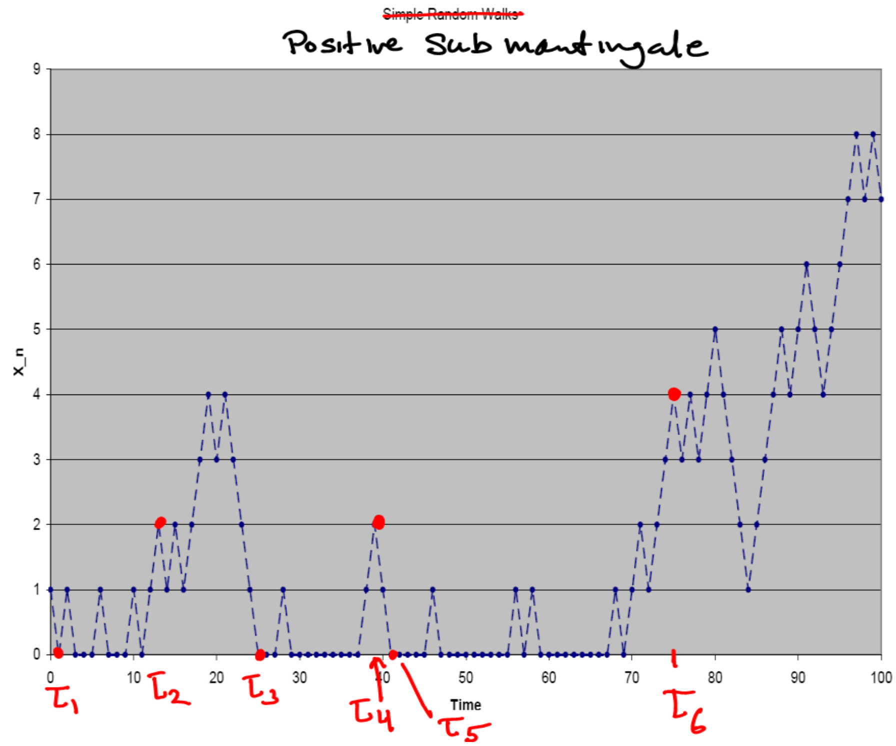


Fig. 19.2. A sample path of a positive submartingale along with stopping times τ_{2j} and τ_{2j+1}, successive hitting times of 2 and 0 respectively. Notice that $X_{\tau_4 \wedge 70} - X_{\tau_3 \wedge 70} \geq 2$ while $X_{\tau_6 \wedge 70} - X_{\tau_5 \wedge 70} \geq 0$. Also observe that $X_{\tau_8 \wedge 90} - X_{\tau_7 \wedge 90} = 0$.

a sequence of stopping times. Suppose that N is given and we choose k such that $2k > N$. Then we know that $\tau_{2k} \geq N$. Thus if we let $\tau_n' := \tau_n \wedge N$, we know that $\tau_n' = N$ for all $n \geq 2k$. Therefore,

$$\begin{aligned} X_N - X_0 &= \sum_{n=1}^{2k} \left(X_{\tau_n'} - X_{\tau_{n-1}'} \right) \\ &= \sum_{n=1}^{k} \left(X_{\tau_{2n}'} - X_{\tau_{2n-1}'} \right) + \sum_{n=1}^{k} \left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} \right) \\ &\geq b U_N^X(0,b) + \sum_{n=1}^{k} \left(X_{\tau_{2n-1}'} - X_{\tau_{2n-2}'} \right), \end{aligned} \tag{19.37}$$

wherein we have used $X_{\tau'_{2n}} - X_{\tau'_{2n-1}} \geq b$ if there were an upcrossing in the interval $\left[\tau'_{2n-1}, \tau'_{2n}\right]$ and $X_{\tau'_{2n}} - X_{\tau'_{2n-1}} \geq 0$ otherwise,[4] see Figure 19.2. Taking expectations of Eq. (19.37) implies

$$\mathbb{E}X_N - \mathbb{E}X_0 \geq b\mathbb{E}U_N^X(0,b) + \sum_{n=1}^{k} \mathbb{E}\left(X_{\tau'_{2n-1}} - X_{\tau'_{2n-2}}\right) \geq b\mathbb{E}U_N^X(0,b)$$

wherein we have used the optional sampling theorem to guarantee,

$$\mathbb{E}\left(X_{\tau'_{2n-1}} - X_{\tau'_{2n-2}}\right) \geq 0.$$

If X is a general submartingale and $-\infty < a < b < \infty$, we know by Jensen's inequality that $(X_n - a)_+$ is still a sub-martingale and moreover

$$U_N^X(a,b) = U^{(X-a)_+}(0, b-a)$$

and therefore

$$\begin{aligned}(b-a)\,\mathbb{E}\left[U_N^X(a,b)\right] &= (b-a)\,\mathbb{E}\left[U^{(X-a)_+}(0,b-a)\right] \\ &\leq \mathbb{E}\,(X_N - a)_+ - \mathbb{E}\,(X_0 - a)_+ .\end{aligned}$$

■

It is worth contemplating a bit how is that $\mathbb{E}\left(X_{\tau'_{2n-1}} - X_{\tau'_{2n-2}}\right) \geq 0$ given that are strategy is to buy high and sell low. On the $\{\tau_{2n-1} \leq N\}$, $X_{\tau_{2n-1}} - X_{\tau_{2n-2}} \leq 0 - b = -b$ and therefore,

$$\begin{aligned}0 \leq& \mathbb{E}\left(X_{\tau'_{2n-1}} - X_{\tau'_{2n-2}}\right) \\ &= \mathbb{E}\left(X_{\tau_{2n-1}} - X_{\tau_{2n-2}} : \tau_{2n-1} \leq N\right) + \mathbb{E}\left(X_{\tau'_{2n-1}} - X_{\tau'_{2n-2}} : \tau_{2n-1} > N\right) \\ \leq& -bP\left(\tau_{2n-1} \leq N\right) + \mathbb{E}\left(X_N - X_{\tau'_{2n-2}} : \tau_{2n-1} > N\right).\end{aligned}$$

Therefore we must have

$$\mathbb{E}\left(X_N - X_{\tau_{2n-2}\wedge N} : \tau_{2n-1} > N\right) \geq bP\left(\tau_{2n-1} \leq N\right)$$

so that X_N must be sufficiently large sufficiently often on the set where $\tau_{2n-1} > N$.

Corollary 19.46. *Suppose* $\{X_n\}_{n=0}^{\infty}$ *is an integrable submartingale such that* $\sup_n \mathbb{E}X_n^+ < \infty$ *(or equivalently* $C := \sup_n \mathbb{E}|X_n| < \infty$, *see Remark 19.18), then* $X_\infty := \lim_{n\to\infty} X_n$ *exists in* $\mathbb{R}$ *a.s. and moreover,* $X_\infty \in L^1(\Omega, \mathcal{B}, P)$.

[4] If $\tau_{2n-1} \geq N$, then $X_{\tau'_{2n}} - X_{\tau'_{2n-1}} = X_N - X_N = 0$, while if $\tau_{2n-1} < N$, $X_{\tau'_{2n}} - X_{\tau'_{2n-1}} = X_{\tau'_{2n}} - 0 \geq 0$.

Proof. For any $-\infty < a < b < \infty$, by Doob's upcrossing inequality (Theorem 19.45) and the MCT,

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \le \frac{1}{b-a}\left[\sup_N \mathbb{E}\left(X_N - a\right)_+ - \mathbb{E}\left(X_0 - a\right)_+\right] < \infty$$

where

$$U_\infty^X(a,b) := \lim_{N\to\infty} U_N^X(a,b)$$

is the total number of upcrossings of X across $[a,b]$. In particular it follows that

$$\Omega_0 := \cap\left\{U_\infty^X(a,b) < \infty : a,b \in \mathbb{Q} \text{ with } a < b\right\}$$

has probability one. Hence by Lemma 19.44, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists in $\bar{\mathbb{R}}$. By Fatou's lemma we know that

$$\mathbb{E}\left[|X_\infty|\right] = \mathbb{E}\left[\liminf_{n\to\infty}|X_n|\right] \le \liminf_{n\to\infty}\mathbb{E}\left[|X_n|\right] \le C < \infty \tag{19.38}$$

and therefore that $X_\infty \in \mathbb{R}$ a.s.

Second Proof. We may also give another proof based on the Krickeberg Decomposition Theorem 19.19 and the supermartingale convergence Corollary 19.54 below. Indeed, by Theorem 19.19, $X_n = M_n - Y_n$ where M is a positive martingale and Y is a positive supermartingale. Hence by two applications of Corollary 19.54 we may conclude that

$$X_\infty = \lim_{n\to\infty} X_n = \lim_{n\to\infty} M_n - \lim_{n\to\infty} Y_n$$

exists in $\mathbb{R}$ almost surely. ∎

Notation 19.47 *Given a probability space, $(\Omega,\mathcal{B},P)$ and $A,B\in\mathcal{B}$, we say $A = B$ a.s. iff $P(A \triangle B) = 0$.*

Corollary 19.48 (Localizing Corollary Eq. 19.46). *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale and $c < \infty$ such that $\Delta_n M \le c$ a.s. for all n. Then*

$$\left\{\lim_{n\to\infty} M_n \text{ exists in } \mathbb{R}\right\} = \left\{\sup_n M_n < \infty\right\} \text{ a.s.}$$

Proof. Let $\tau_m := \inf\{n : M_n \ge m\}$ for all $m \in \mathbb{N}$. Then by the optional stopping theorem, $n \to M_n^{\tau_m}$ is still a martingale. Since $M_n^{\tau_m} \le m + c$, it follows that $\mathbb{E}\left(M_n^{\tau_m}\right)_+ \le m + c < \infty$ for all n. Hence we may apply Corollary 19.46 to conclude, $\lim_{n\to\infty} M_n^{\tau_m} = M_\infty^{\tau_m}$ exists in $\mathbb{R}$ almost surely. Therefore $n \to M_n$ is convergent in $\mathbb{R}$ almost surely on the set

$$\cup_m\left\{M^{\tau_m} = M\right\} = \left\{\sup_n M_n < \infty\right\}.$$

Conversely if $n \to M_n$ is convergent in $\mathbb{R}$, then $\sup_n M_n < \infty$. ∎

Corollary 19.49. *Suppose $M = \{M_n\}_{n=0}^{\infty}$ is a martingale, and $c < \infty$ such that $|\Delta_n M| \le c$ a.s. for all n. Let*

$$C := \left\{ \lim_{n\to\infty} M_n \text{ exists in } \mathbb{R} \right\} \text{ and}$$
$$D := \left\{ \limsup_{n\to\infty} M_n = \infty \right\} \cap \left\{ \liminf_{n\to\infty} M_n = -\infty \right\}.$$

Then, $P(C \cup D) = 1$.

Proof. Since both M and $-M$ satisfy the hypothesis of Corollary 19.48, we may conclude that almost surely,

$$C = \left\{ \sup_n M_n < \infty \right\} = \left\{ \inf_n M_n > -\infty \right\}$$

and hence almost surely,

$$C^c = \left\{ \sup_n M_n = \infty \right\} = \left\{ \inf_n M_n = -\infty \right\}$$
$$= \left\{ \sup_n M_n = \infty \right\} \cap \left\{ \inf_n M_n = -\infty \right\} = D.$$

■

Corollary 19.50. *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ is a filtered probability space and $A_n \in \mathcal{B}_n$ for all n. Then*

$$\{A_n \text{ i.o.}\} = \left\{ \sum_n \mathbb{E}\left[1_{A_n} | \mathcal{B}_{n-1}\right] = \infty \right\} \text{ a.s.} \tag{19.39}$$

Proof. Let $\Delta_n M := 1_{A_n} - \mathbb{E}\left[1_{A_n} | \mathcal{B}_{n-1}\right]$ and then set $M_n := \sum_{k\le n} \Delta_n M$. Then M is a martingale with $|\Delta_n M| \le 1$ for all n. Since

$$\{A_n \text{ i.o.}\} = \left\{ \sum_n 1_{A_n} = \infty \right\},$$

it follows that on C we have $\{A_n \text{ i.o.}\} = \left\{\sum_n \mathbb{E}\left[1_{A_n} | \mathcal{B}_{n-1}\right] = \infty\right\}$ a.s. Moreover, on D, we must have $\sum_n 1_{A_n} = \infty$ and $\sum_n \mathbb{E}\left[1_{A_n} | \mathcal{B}_{n-1}\right] = \infty$ and hence again it follows that Eq. (19.39) holds. Since $C \cup D = \Omega$ a.s., the proof is complete. ■

See Durrett [**?**, Chapter 4.3] for more in this direction.

19.6 Supermartingale inequalities

As the optional sampling theorem was our basic tool for deriving submartingale inequalities, the following switching lemma will be our basic tool for deriving positive supermartingale inequalities.

Lemma 19.51 (Switching Lemma). *Suppose that X and Y are two supermartingales and τ is a stopping time such that $X_\tau \geq Y_\tau$ on $\{\tau < \infty\}$. Then*

$$Z_n = 1_{n<\tau} X_n + 1_{n \geq \tau} Y_n = \begin{cases} X_n & \text{if } n < \tau \\ Y_n & \text{if } n \geq \tau \end{cases}$$

is again a supermartingale. (In short we can switch from X to Y at time, τ, provided $Y \leq X$ at the switching time, τ.) This lemma is valid if $X_n, Y_n \in L^1(\Omega, \mathcal{B}_n, P)$ for all n or if both $X_n, Y_n \geq 0$ for all n. In the latter case, we should be using the extended notion of conditional expectations.

Proof. We begin by observing,

$$\begin{aligned} Z_{n+1} &= 1_{n+1<\tau} X_{n+1} + 1_{n+1 \geq v} Y_{n+1} \\ &= 1_{n+1<\tau} X_{n+1} + 1_{n \geq v} Y_{n+1} + 1_{\tau = n+1} Y_{n+1} \\ &\leq 1_{n+1<\tau} X_{n+1} + 1_{n \geq v} Y_{n+1} + 1_{\tau = n+1} X_{n+1} \\ &= 1_{n<\tau} X_{n+1} + 1_{n \geq v} Y_{n+1}. \end{aligned}$$

Since $\{n < \tau\}$ and $\{n \geq \tau\}$ are $\mathcal{B}_n$ – measurable, it now follows from the supermartingale property of X and Y that

$$\begin{aligned} \mathbb{E}_{\mathcal{B}_n} Z_{n+1} &\leq \mathbb{E}_{\mathcal{B}_n} \left[1_{n<\tau} X_{n+1} + 1_{n \geq v} Y_{n+1}\right] \\ &= 1_{n<\tau} \mathbb{E}_{\mathcal{B}_n} \left[X_{n+1}\right] + 1_{n \geq v} \mathbb{E}_{\mathcal{B}_n} \left[Y_{n+1}\right] \\ &\leq 1_{n<\tau} X_n + 1_{n \geq v} Y_n = Z_n. \end{aligned}$$

■

19.6.1 Maximal Inequalities

Theorem 19.52 (Supermartingale maximal inequality). *Let X be a positive supermartingale (in the extended sense) and $a \in \mathcal{B}_0$ with $a \geq 0$, then*

$$aP\left[\sup_n X_n \geq a | \mathcal{B}_0\right] \leq a \wedge X_0 \tag{19.40}$$

and moreover

$$P\left[\sup_n X_n = \infty | \mathcal{B}_0\right] = 0 \text{ on } \{X_0 < \infty\}. \tag{19.41}$$

In particular if $X_0 < \infty$ a.s. then $\sup_n X_n < \infty$ a.s.

Proof. Let $\tau := \inf\{n : X_n \geq a\}$ which is a stopping time since,

$$\{\tau \leq n\} = \{X_n \geq a\} \in \mathcal{B}_n \text{ for all } n.$$

Since $X_\tau \geq a$ on $\{\tau < \infty\}$ and $Y_n := a$ is a supermartingale, it follows by the switching Lemma 19.51 that

$$Z_n := 1_{n<\tau} X_n + a 1_{n \geq \tau}$$

is a supermartingale (in the extended sense). In particular it follows

$$aP(\tau \leq n|\mathcal{B}_0) = \mathbb{E}_{\mathcal{B}_0}[a 1_{n\geq\tau}] \leq \mathbb{E}_{\mathcal{B}_0} Z_n \leq Z_0,$$

and

$$Z_0 = 1_{0<\tau}\ X_0 + a 1_{\tau=0} = 1_{X_0<a} X_0 + 1_{X_0 \geq a} a = a \wedge X_0.$$

Therefore, using the cMCT,

$$\begin{aligned} aP\left[\sup_n X_n \geq a|\mathcal{B}_0\right] &= aP[\tau < \infty|\mathcal{B}_0] = \lim_{n\to\infty} aP(\tau \leq n|\mathcal{B}_0) \\ &\leq Z_0 = a \wedge X_0 \end{aligned}$$

which proves Eq. (19.40).

For the last assertion, take $a > 0$ to be constant in Eq. (19.40) and then use the cDCT to let $a \uparrow \infty$ to conclude

$$P\left[\sup_n X_n = \infty|\mathcal{B}_0\right] = \lim_{a\uparrow\infty} P\left[\sup_n X_n \geq a|\mathcal{B}_0\right] \leq \lim_{a\uparrow\infty} 1 \wedge \frac{X_0}{a} = 1_{X_0=\infty}.$$

Multiplying this equation by $1_{X_0<\infty}$ and then taking expectations implies

$$\mathbb{E}\left[1_{\sup_n X_n=\infty} 1_{X_0<\infty}\right] = \mathbb{E}\left[1_{X_0=\infty} 1_{X_0<\infty}\right] = 0$$

which implies $1_{\sup_n X_n=\infty} 1_{X_0<\infty} = 0$ a.s., i.e. $\sup_n X_n < \infty$ a.s. on $\{X_0 < \infty\}$. ■

19.6.2 The upcrossing inequality and convergence result

Theorem 19.53 (Dubin's Upcrossing Inequality). *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a positive supermartingale and $0 < a < b < \infty$. Then*

$$P\left(U_\infty^X(a,b) \geq k|\mathcal{B}_0\right) \leq \left(\frac{a}{b}\right)^k \left(1 \wedge \frac{X_0}{a}\right), \text{ for } k \geq 1 \tag{19.42}$$

and $U_\infty(a,b) < \infty$ a.s. and in fact

$$\mathbb{E}\left[U_\infty^X(a,b)\right] \leq \frac{1}{b/a - 1} = \frac{a}{b-a} < \infty.$$

Proof. Since

$$U_N^X(a,b) = U_N^{X/a}(1, b/a),$$

it suffices to consider the case where $a = 1$ and $b > 1$. Let τ_n be the stopping times defined in Eq. (19.36) with $a = 1$ and $b > 1$, i.e.

$$\tau_0 = 0,\ \tau_1 = \inf\{n \geq \tau_0 : X_n \leq 1\}$$
$$\tau_2 = \inf\{n \geq \tau_1 : X_n \geq b\},\ \tau_3 := \inf\{n \geq \tau_2 : X_n \leq 1\}$$
$$\vdots$$
$$\tau_{2k} = \inf\{n \geq \tau_{2k-1} : X_n \geq b\},\ \tau_{2k+1} := \inf\{n \geq \tau_{2k} : X_n \leq 1\},$$
$$\vdots$$

see Figure 19.1.

Let $k \geq 1$ and use the switching Lemma 19.51 repeatedly to define a new positive supermatingale $Y_n = Y_n^{(k)}$ (see Exercise 19.5 below) as follows,

$$\begin{aligned} Y_n^{(k)} &= 1_{n<\tau_1} + 1_{\tau_1 \leq n < \tau_2} X_n \\ &+ b1_{\tau_2 \leq n < \tau_3} + bX_n 1_{\tau_3 \leq n < \tau_4} \\ &+ b^2 1_{\tau_4 \leq n < \tau_5} + b^2 X_n 1_{\tau_5 \leq n < \tau_6} \\ &\vdots \\ &+ b^{k-1} 1_{\tau_{2k-2} \leq n < \tau_{2k-1}} + b^{k-1} X_n 1_{\tau_{2k-1} \leq n < \tau_{2k}} \\ &+ b^k 1_{\tau_{2k} \leq n}. \end{aligned} \tag{19.43}$$

Since $\mathbb{E}[Y_n|\mathcal{B}_0] \leq Y_0$ a.s., $Y_n \geq b^k 1_{\tau_{2k} \leq n}$, and

$$Y_0 = 1_{0<\tau_1} + 1_{\tau_1=0} X_0 = 1_{X_0>1} + 1_{X_0 \leq 1} X_0 = 1 \wedge X_0,$$

we may infer that

$$b^k P(\tau_{2k} \leq n|\mathcal{B}_0) = \mathbb{E}\left[b^k 1_{\tau_{2k} \leq n}|\mathcal{B}_0\right] \leq \mathbb{E}[Y_n|\mathcal{B}_0] \leq 1 \wedge X_0 \text{ a.s.}$$

Using $cMCT$, we may now let $n \to \infty$ to conclude

$$P\left(U^X(1,b) \geq k|\mathcal{B}_0\right) \leq P(\tau_{2k} < \infty|\mathcal{B}_0) \leq \frac{1}{b^k}(1 \wedge X_0) \text{ a.s.}$$

which is Eq. (19.42). Using cDCT, we may let $k \uparrow \infty$ in this equation to discover $P\left(U_\infty^X(1,b) = \infty|\mathcal{B}_0\right) = 0$ a.s. and in particular, $U_\infty^X(1,b) < \infty$ a.s. In fact we have

$$\begin{aligned} \mathbb{E}\left[U_\infty^X(1,b)\right] &= \sum_{k=1}^\infty P\left(U_\infty^X(1,b) \geq k\right) \leq \sum_{k=1}^\infty \mathbb{E}\left[\frac{1}{b^k}(1 \wedge X_0)\right] \\ &= \frac{1}{b}\frac{1}{1-1/b}\mathbb{E}[(1 \wedge X_0)] \leq \frac{1}{b-1} < \infty. \end{aligned}$$

■

Exercise 19.5. In this exercise you are asked to fill in the details showing Y_n in Eq. (19.43) is still a supermartingale. To do this, define $Y_n^{(k)}$ via Eq. (19.43) and then show (making use of the switching Lemma 19.51twice) $Y_n^{(k+1)}$ is a submartingale under the assumption that $Y_n^{(k)}$ is a submartingale. Finish off the induction argument by observing that the constant process, $U_n := 1$ and $V_n = 0$ are supermartingales such that $U_{\tau_1} = 1 \geq 0 = V_{\tau_1}$ on $\{\tau_1 < \infty\}$, and therefore by the switching Lemma 19.51,

$$Y_n^{(1)} = 1_{0 \leq n < \tau_1} U_n + 1_{\tau_1 \leq n} V_n = 1_{0 \leq n < \tau_1}$$

is also a supermartingale.

Corollary 19.54 (Positive Supermartingale convergence). *Suppose $X = \{X_n\}_{n=0}^{\infty}$ is a positive supermartingale (possibly in the extended sense), then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and we have*

$$\mathbb{E}\left[X_\infty|\mathcal{B}_n\right] \leq X_n \text{ for all } n \in \bar{\mathbb{N}}. \tag{19.44}$$

In particular,

$$\mathbb{E}X_\infty \leq \mathbb{E}X_n \leq \mathbb{E}X_0 \text{ for all } n < \infty. \tag{19.45}$$

Proof. The set,

$$\Omega_0 := \cap \left\{U_\infty^X(a,b) < \infty : a, b \in \mathbb{Q} \text{ with } a < b\right\},$$

has full measure ($P(\Omega_0) = 1$) by Dubin's upcrossing inequality in Theorem 19.53. So by Lemma 19.44, for $\omega \in \Omega_0$ we have $X_\infty(\omega) := \lim_{n\to\infty} X_n(\omega)$ exists[5] in $[0, \infty]$. For definiteness, let $X_\infty = 0$ on Ω_0^c. Equation (19.44) is now a consequence of cFatou;

$$\mathbb{E}\left[X_\infty|\mathcal{B}_n\right] = \mathbb{E}\left[\lim_{m\to\infty} X_m|\mathcal{B}_n\right] \leq \liminf_{m\to\infty} \mathbb{E}\left[X_m|\mathcal{B}_n\right] \leq \liminf_{m\to\infty} X_n = X_n \text{ a.s.}$$

The supermartingale property guarantees that $\mathbb{E}X_n \leq \mathbb{E}X_0$ for all $n < \infty$ while taking expectations of Eq. (19.44) implies $\mathbb{E}X_\infty \leq \mathbb{E}X_n$. ■

Theorem 19.55 (Optional sampling II – Positive supermartingales). *Suppose that $X = \{X_n\}_{n=0}^{\infty}$ is a positive supermartingale, $X_\infty := \lim_{n\to\infty} X_n$ (which exists a.s. by Corollary 19.54), and σ and τ are **arbitrary** stopping times. Then $X_n^\tau := X_{\tau\wedge n}$ is a positive $\{\mathcal{B}_n\}_{n=0}^{\infty}$ – super martingale, $X_\infty^\tau = \lim_{n\to\infty} X_{\tau\wedge n}^\tau$, and*

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] \leq X_{\sigma\wedge\tau} \text{ a.s.} \tag{19.46}$$

Moreover, if $\mathbb{E}X_0 < \infty$, then $\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[X_\infty^\tau\right] < \infty$.

Proof. We already know that X^τ is a positive supermatingale by optional stopping Theorem 19.35. Hence an application of Corollary 19.54 implies that $\lim_{n\to\infty} X_n^\tau = \lim_{n\to\infty} X_{\tau\wedge n}$ is convergent and

[5] If $\mathbb{E}X_0 < \infty$, this may also be deduced by applying Corollary 19.46 to $\{-X_n\}_{n=0}^{\infty}$.

$$\mathbb{E}\left[\lim_{n\to\infty} X_n^\tau|\mathcal{B}_m\right] \le X_m^\tau = X_{\tau\wedge m} \text{ for all } m < \infty. \tag{19.47}$$

On the set $\{\tau < \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ and on the set $\{\tau = \infty\}$, $\lim_{n\to\infty} X_{\tau\wedge n} = \lim_{n\to\infty} X_n = X_\infty = X_\tau$ a.s. Therefore it follows that $\lim_{n\to\infty} X_n^\tau = X_\tau$ and Eq. (19.47) may be expressed as

$$\mathbb{E}\left[X_\tau|\mathcal{B}_m\right] \le X_{\tau\wedge m} \text{ for all } m < \infty. \tag{19.48}$$

An application of Lemma 19.29 now implies

$$\mathbb{E}\left[X_\tau|\mathcal{B}_\sigma\right] = \sum_{m\le\infty} 1_{\sigma=m}\mathbb{E}\left[X_\tau|\mathcal{B}_m\right] \le \sum_{m\le\infty} 1_{\sigma=m} X_{\tau\wedge m} = X_{\tau\wedge\sigma} \text{ a.s.}$$

■

19.7 Martingale Closure and Regularity Results

Theorem 19.56. *Let* $M := \{M_n\}_{n=0}^\infty$ *be an* L^1 *– bounded martingale, i.e.* $C := \sup_n \mathbb{E}|M_n| < \infty$ *and let* $M_\infty := \lim_{n\to\infty} M_n$ *which exists a.s. and satisfies,* $\mathbb{E}|M_\infty| < \infty$ *by Corollary 19.46. Then the following are equivalent;*

1. *There exists* $X \in L^1(\Omega, \mathcal{B}, P)$ *such that* $M_n = \mathbb{E}[X|\mathcal{B}_n]$ *for all* n.
2. $\{M_n\}_{n=0}^\infty$ *is uniformly integrable.*
3. $M_n \to M_\infty$ *in* $L^1(\Omega, \mathcal{B}, P)$.

Moreover, if any of the above equivalent conditions hold we may take $X = M_\infty$.

Proof. 1. $\implies$ 2. was already proved in Proposition 19.7. 2. $\implies$ 3. follows from Theorem 11.31.

3. $\implies$ 2. If $M_n \to M_\infty$ in $L^1(\Omega, \mathcal{B}, P)$ and $A \in \mathcal{B}_m$, then $\mathbb{E}[M_n : A] = \mathbb{E}[M_m : A]$ for all $n \ge m$ and

$$\mathbb{E}[M_\infty : A] = \lim_{n\to\infty} \mathbb{E}[M_n : A] = \mathbb{E}[M_m : A].$$

Since $A \in \mathcal{B}_m$ was arbitrary, it follows that $M_n = \mathbb{E}[M_\infty|\mathcal{B}_n]$. ■

Definition 19.57. *A martingale satisfying any and all of the equivalent statements in Theorem 19.56 is said to be* ***regular.***

Theorem 19.58. *Suppose* $1 < p < \infty$ *and* $M := \{M_n\}_{n=0}^\infty$ *is an* L^p *– bounded martingale. Then* $M_n \to M_\infty$ *almost surely and in* L^p.

Proof. Again, the almost sure convergence follows from Corollary 19.46. So, because of Corollary 11.34, to finish the proof it suffices to show $\{|M_n|^p\}_{n=0}^\infty$ is uniformly integrable. But by Doob's inequality, Corollary 19.43, and the MCT, we find

$$\mathbb{E}\left[\sup_n |M_n|^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}\left[|M_n|^p\right] < \infty.$$

It now follows by an application of Proposition 11.29 that $\{|M_n|^p\}_{n=0}^\infty$ is uniformly integrable. ■

Theorem 19.59 (Optional sampling III – regular martingales). *Suppose that $M = \{M_n\}_{n=0}^\infty$ is a regular martingale, σ and τ are **arbitrary** stopping times. Define $M_\infty := \lim_{n\to\infty} M_n$ a.s.. Then $M_\infty \in L^1(P)$,*

$$M_\tau = \mathbb{E}\left[M_\infty|\mathcal{B}_\tau\right], \ \mathbb{E}|M_\tau| < \infty \tag{19.49}$$

and

$$\mathbb{E}\left[M_\tau|\mathcal{B}_\sigma\right] = M_{\sigma\wedge\tau} \ \text{a.s.} \tag{19.50}$$

Proof. By Theorem 19.56, $M_\infty \in L^1(\Omega, \mathcal{B}, P)$ and $M_n := \mathbb{E}_{\mathcal{B}_n} M_\infty$ a.s. for all $n \le \infty$. By Lemma 19.29,

$$\mathbb{E}_{\mathcal{B}_\tau} M_\infty = \sum_{n\le\infty} 1_{\tau=n} \mathbb{E}_{\mathcal{B}_n} M_\infty = \sum_{n\le\infty} 1_{\tau=n} M_n = M_\tau.$$

Hence we have $|M_\tau| = |\mathbb{E}_{\mathcal{B}_\tau} M_\infty| \le \mathbb{E}_{\mathcal{B}_\tau}|M_\infty|$ a.s. and $\mathbb{E}|M_\tau| \le \mathbb{E}|M_\infty| < \infty$. An application of Exercise 19.3 now concludes the proof;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \mathbb{E}_{\mathcal{B}_\sigma}\mathbb{E}_{\mathcal{B}_\tau} M_\infty = \mathbb{E}_{\mathcal{B}_{\sigma\wedge\tau}} M_\infty = M_{\sigma\wedge\tau}.$$

■

Definition 19.60. *Let $M = \{M_n\}_{n=0}^\infty$ be a martingale. We say that τ is a **regular stopping time for** M if M^τ is a regular martingale.*

Remark 19.61. If τ is regular for M, then $\lim_{n\to\infty} M_n^\tau := M_\infty^\tau$ exists a.s. and hence

$$\lim_{n\to\infty} M_n = M_\infty^\tau \text{ a.s. on } \{\tau = \infty\}. \tag{19.51}$$

Thus if τ is regular of M, we may define M_τ as,

$$M_\tau := M_\infty^\tau = \lim_{n\to\infty} M_{n\wedge\tau}.$$

Also observe by Fatou's lemma that,

$$\mathbb{E}|M_\tau| \le \liminf_{n\to\infty} \mathbb{E}|M_n^\tau| \le \sup_n \mathbb{E}|M_n^\tau|.$$

Theorem 19.62. *Suppose $M = \{M_n\}_{n=0}^\infty$ is a martingale and σ, τ, are stopping times such that τ is a regular stopping time for M. Then*

1.

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = M_{\tau\wedge\sigma}. \tag{19.52}$$

2. If $\sigma \le \tau$ a.s. then $M_n^\sigma = \mathbb{E}_{\mathcal{B}_n}[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau]$ and σ is regular for M.

Proof. By assumption, $M_\tau = \lim_{n\to\infty} M_{n\wedge\tau}$ exists almost surely and in $L^1(P)$ and $M_n^\tau = \mathbb{E}[M_\tau|\mathcal{B}_n]$ for $n \le \infty$.

1. Equation (19.52) is a consequence of;

$$\mathbb{E}_{\mathcal{B}_\sigma} M_\tau = \sum_{n\le\infty} 1_{\sigma=n}\mathbb{E}_{\mathcal{B}_n} M_\tau = \sum_{n\le\infty} 1_{\sigma=n} M_n^\tau = M_{\sigma\wedge\tau}.$$

2. By Theorem 19.59 and Exercise 19.3,

$$M_n^\sigma = M_{\sigma\wedge n} = M_{\sigma\wedge n}^\tau = \mathbb{E}_{\mathcal{B}_{\sigma\wedge n}} M_\infty^\tau = \mathbb{E}_{\mathcal{B}_{\sigma\wedge n}} M_\tau = \mathbb{E}_{\mathcal{B}_n}[\mathbb{E}_{\mathcal{B}_\sigma} M_\tau]$$

from which it follows that M^σ is a regular martingale. ■

Proposition 19.63. *Suppose that M is a martingale and τ is a stopping time. Then the τ is regular for M iff;*

1. $\mathbb{E}[|M_\tau| : \tau < \infty] < \infty$ and
2. $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is a uniformly integrable sequence of random variables.

Moreover, condition 1. is automatically satisfied if M is L^1 – bounded, i.e. if $C := \sup_n \mathbb{E}|M_n| < \infty$.

Proof. ($\Longrightarrow$) If τ is regular for M, $M_\tau \in L^1(P)$ and $M_n = \mathbb{E}_{\mathcal{B}_n} M_\tau$. In particular it follows that

$$\mathbb{E}[|M_\tau| : \tau < \infty] \le \mathbb{E}|M_\tau| < \infty.$$

Moreover,

$$|M_n 1_{n<\tau}| \le |\mathbb{E}_{\mathcal{B}_n} M_\tau 1_{n<\tau}| \le \mathbb{E}_{\mathcal{B}_n}|M_\tau| \text{ a.s.}$$

from which it follows that $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is uniformly integrable.

($\Longleftarrow$) Our goal is to show $\{M_n^\tau\}_{n=0}^\infty$ is uniformly integrable. We begin with the identity;

$$\begin{aligned}\mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a] &= \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ \tau \le n] \\ &\quad + \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau].\end{aligned}$$

Since (by assumption 1.) $\mathbb{E}[|M_\tau 1_{\tau<\infty}|] < \infty$ and

$$\mathbb{E}[|M_\tau| : |M_\tau| \ge a,\ \tau \le n] \le \mathbb{E}[|M_\tau 1_{\tau<\infty}| : |M_\tau 1_{\tau<\infty}| \ge a],$$

if follows that

$$\lim_{a\to\infty}\sup_n \mathbb{E}[|M_\tau| : |M_\tau| \ge a,\ \tau \le n] = 0.$$

Moreover,

$$\sup_n \mathbb{E}[|M_n^\tau| : |M_n^\tau| \ge a,\ n < \tau] = \sup_n \mathbb{E}[|M_n^\tau 1_{n<\tau}| : |M_n^\tau 1_{n<\tau}| \ge a]$$

goes to zero as $n \to \infty$ by assumption 2. Hence we have shown,

$$\lim_{a\to\infty} \sup_n \mathbb{E}\left[|M_n^\tau| : |M_n^\tau| \geq a\right] = 0$$

as desired.

For the last assertion, by Corollary 19.46, $M_\infty := \lim_{n\to\infty} M_n$ a.s. and $\mathbb{E}\,|M_\infty| < \infty$. Therefore,

$$\mathbb{E}\left[|M_\tau| : \tau < \infty\right] \leq \mathbb{E}\,|M_\tau| = \mathbb{E}\left[\lim_{n\to\infty} |M_{\tau\wedge n}|\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[|M_{\tau\wedge n}|\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[|M_n|\right] < \infty$$

wherein the second to last inequality we have used the optional sampling theorem to conclude

$$|M_{\tau\wedge n}| \leq \mathbb{E}_{\mathcal{B}_{\tau\wedge n}} |M_n|\,.$$

■

Corollary 19.64. *Suppose that M is an L^1 – bounded martingale and $J \in \mathcal{B}_{\mathbb{R}}$ is a bounded set, then $\tau = \inf\{n : M_n \notin J\}$ is a regular stopping time for M.*

Proof. According to Proposition 19.63, it suffices to show $\{M_n 1_{n<\tau}\}_{n=0}^\infty$ is a uniformly integrable sequence of random variables. However, if we choose $A < \infty$ such that $J \subset [-A, A]$, since $M_n 1_{n<\tau} \in J$ we have $|M_n 1_{n<\tau}| \leq A$ which is sufficient to complete the proof. ■

For the next three exercises, let $\{Z_n\}_{n=1}^\infty$ be a sequence of Bernoulli random variables with $P(Z_n = \pm 1) = \frac{1}{2}$ and let $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$. Then S becomes a martingale relative to the filtration, $\mathcal{B}_n := \sigma(Z_1, \ldots, X_n)$ with $\mathcal{B}_0 := \{\emptyset, \Omega\}$ – of course S_n is the (fair) simple random walk on $\mathbb{Z}$. For any $a \in \mathbb{Z}$, let

$$\sigma_a := \inf\{n : S_n = a\}\,.$$

Exercise 19.6. For $a < 0 < b$ with $a, b \in \mathbb{Z}$, let $\tau = \sigma_a \wedge \sigma_b$. Explain why τ is regular for S. Use this to show $P(\tau = \infty) = 0$. **Hint:** make use of Remark 19.61 and the fact that $|S_n - S_{n-1}| = |Z_n| = 1$ for all n.

Exercise 19.7. In this exercise, you are asked to give a central limit theorem argument for the result, $P(\tau = \infty) = 0$, Exercise 19.6. **Hints:** Use the central limit theorem to show

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\, e^{-x^2/2} dx \geq f(0)\, P(\tau = \infty) \tag{19.53}$$

for all bounded continuous functions, $f : \mathbb{R} \to [0, \infty)$. Use this inequality to conclude that $P(\tau = \infty) = 0$.

Exercise 19.8. Show

$$P(\sigma_b < \sigma_a) = \frac{|a|}{b + |a|} \tag{19.54}$$

and use this to conclude $P(\sigma_b < \infty) = 1$, i.e. every $b \in \mathbb{N}$ is almost surely visited by S_n. (This last result also follows by the Hewitt-Savage Zero-One Law, see Example 7.45 where it is shown b is visited infinitely often.)

Hint: Using properties of martingales and Exercise 19.6, compute $\lim_{n\to\infty} \mathbb{E}[S_n^{\sigma_a \wedge \sigma_b}]$ in two different ways.

Exercise 19.9. Let $\{Z_n\}_{n=1}^{\infty}$ be independent random variables, $S_0 = 0$ and $S_n := Z_1 + \cdots + Z_n$, and $f_n(\lambda) := \mathbb{E}\left[e^{i\lambda Z_n}\right]$. Suppose $\mathbb{E}e^{i\lambda S_n} = \prod_{n=1}^{N} f_n(\lambda)$ converges to a continuous function, $F(\lambda)$, as $N \to \infty$. Show for each $\lambda \in \mathbb{R}$ that

$$P\left(\lim_{n\to\infty} e^{i\lambda S_n} \text{ exists}\right) = 1. \tag{19.55}$$

Hints:

1. Show it is enough to find an $\varepsilon > 0$ such that Eq. (19.55) holds for $|\lambda| \le \varepsilon$.
2. Choose $\varepsilon > 0$ such that $|F(\lambda) - 1| < 1/2$ for $|\lambda| \le \varepsilon$. For $|\lambda| \le \varepsilon$, show $M_n(\lambda) := \frac{e^{i\lambda S_n}}{\mathbb{E}e^{i\lambda S_n}}$ is a bounded complex[6] martingale relative to the filtration, $\mathcal{B}_n = \sigma(Z_1, \ldots, Z_n)$.

Exercise 19.10 (Continuation of Exercise 19.9). Let $\{Z_n\}_{n=1}^{\infty}$ be independent random variables. Prove the series, $\sum_{n=1}^{\infty} Z_n$, converges in $\mathbb{R}$ a.s. iff $\prod_{n=1}^{N} f_n(\lambda)$ converges to a continuous function, $F(\lambda)$ as $N \to \infty$. Conclude from this that $\sum_{n=1}^{\infty} Z_n$ is a.s. convergent iff $\sum_{n=1}^{\infty} Z_n$ is convergent in distribution. (See Doob [**?**, Chapter VII.5].)

19.8 Backwards Submartingales

In this section we will consider submartingales indexed by $\mathbb{Z}_- := \{\ldots, -n, -n+1, \ldots, -2, -1, 0\}$. So again we assume that we have an increasing filtration, $\{\mathcal{B}_n : n \le 0\}$, i.e. $\cdots \subset \mathcal{B}_{-2} \subset \mathcal{B}_{-1} \subset \mathcal{B}_0 \subset \mathcal{B}$. As usual, we say an adapted process $\{X_n\}_{n\le 0}$ is a submartingale (martingale) provided $\mathbb{E}[X_m - X_n|\mathcal{B}_n] \ge 0$ $(=0)$ for all $m \ge n$. Observe that $\mathbb{E}X_m \ge \mathbb{E}X_n$ for $m \ge n$, so that $\mathbb{E}X_{-n}$ decreases as n increases. Also observe that $\left(X_{-n}, X_{-(n-1)}, \ldots, X_{-1}, X_0\right)$ is a "finite string" submartingale relative to the filtration, $\mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \cdots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0$.

Theorem 19.65 (Backwards Submartingale Convergence). *Let $\{\mathcal{B}_n : n \le 0\}$ be a reverse filtration, $\{X_n\}_{n\le 0}$ is a backwards submartingale. Then $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. in $\{-\infty\} \cup \mathbb{R}$ and $X_{-\infty}^{+} \in L^1(\Omega, \mathcal{B}, P)$. If we further assume that*

$$C := \lim_{n\to-\infty} \mathbb{E}X_n = \inf_{n\le 0} \mathbb{E}X_n > -\infty, \tag{19.56}$$

then $\{X_n\}_{n\le 0}$ uniformly integrability, $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$, and $\lim_{n\to-\infty} \mathbb{E}|X_n - X_{-\infty}| = 0$.

[6] Please use the obvious generalization of a martingale for complex valued processes. It will be useful to observe that the real and imaginary parts of a complex martingales are real martingales.

Proof. The number of downcrossings of $\left(X_0, X_{-1}, \dots, X_{-(n-1)}, X_{-n}\right)$ across $[a,b]$, (denoted by $D_n(a,b)$) is equal to the number of upcrossings, $\left(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0\right)$ across $[a,b]$. Since $\left(X_{-n}, X_{-(n-1)}, \dots, X_{-1}, X_0\right)$ is a $\mathcal{B}_{-n} \subset \mathcal{B}_{-(n-1)} \subset \dots \subset \mathcal{B}_{-1} \subset \mathcal{B}_0$ submartingale, we may apply Doob's upcrossing inequality (Theorem 19.45) to find;

$$\begin{aligned}(b-a)\,\mathbb{E}\left[D_n(a,b)\right] &\leq \mathbb{E}\left(X_0 - a\right)_+ - \mathbb{E}\left(X_{-n} - a\right)_+ \\ &\leq \mathbb{E}\left(X_0 - a\right)_+ < \infty. \qquad (19.57)\end{aligned}$$

Letting $D_\infty(a,b) :=\uparrow \lim_{n\to\infty} D_n(a,b)$ be the total number of downcrossing of $\left(X_0, X_{-1}, \dots, X_{-n}, \dots\right)$, using the MCT to pass to the limit in Eq. (19.57), we have

$$(b-a)\,\mathbb{E}\left[D_\infty(a,b)\right] \leq \mathbb{E}\left(X_0 - a\right)_+ < \infty.$$

In particular it follows that $D_\infty(a,b) < \infty$ a.s. for all $a < b$.

As in the proof of Corollary 19.46 (making use of the obvious downcrossing analogue of Lemma 19.44), it follows that $X_{-\infty} := \lim_{n\to-\infty} X_n$ exists in $\bar{\mathbb{R}}$ a.s. At the end of the proof, we will show that $X_{-\infty}$ takes values in $\{-\infty\} \cup \mathbb{R}$ almost surely.

Now suppose that $C > -\infty$. We begin by computing the Doob decomposition of X_n as $X_n = M_n + A_n$ with A_n being predictable, increasing and satisfying, $A_{-\infty} = \lim_{n\to-\infty} A_n = 0$. If such an A is to exist, following Lemma 19.15, we should define

$$A_n = \sum_{k\leq n} \mathbb{E}\left[\Delta_k X | \mathcal{B}_{k-1}\right].$$

This is a well defined increasing predictable process since that submartingale property implies $\mathbb{E}\left[\Delta_k X | \mathcal{B}_{k-1}\right] \geq 0$. Moreover we have

$$\begin{aligned}\mathbb{E}A_0 &= \sum_{k\leq 0} \mathbb{E}\left[\mathbb{E}\left[\Delta_k X | \mathcal{B}_{k-1}\right]\right] = \sum_{k\leq 0} \mathbb{E}\left[\Delta_k X\right] \\ &= \lim_{N\to\infty}\left(\mathbb{E}X_0 - \mathbb{E}X_{-N}\right) = \mathbb{E}X_0 - \inf_{n\leq 0} \mathbb{E}X_n = \mathbb{E}X_0 - C < \infty.\end{aligned}$$

As $0 \leq A_n \leq A_n^* = A_0 \in L^1(P)$, it follows that $\{A_n\}_{n\leq 0}$ is uniformly integrable Moreover if we define $M_n := X_n - A_n$, then

$$\mathbb{E}\left[\Delta_n M | \mathcal{B}_{n-1}\right] = \mathbb{E}\left[\Delta_n X - \Delta_n A | \mathcal{B}_{n-1}\right] = \mathbb{E}\left[\Delta_n X | \mathcal{B}_{n-1}\right] - \Delta_n A = 0 \text{ a.s.}$$

Thus M is a martingale and therefore, $M_n = \mathbb{E}\left[M_0 | \mathcal{B}_n\right]$ with $M_0 = X_0 - A_0 \in L^1(P)$. An application of Proposition 19.7 implies $\{M_n\}_{n\leq 0}$ is uniformly integrable and hence $X_n = M_n + A_n$ is uniformly integrable as well. (See Remark 19.66 for an alternate proof of the uniform integrability of X.)

Therefore $X_{-\infty} \in L^1(\Omega, \mathcal{B}, P)$ and $X_n \to X_{-\infty}$ in $L^1(\Omega, \mathcal{B}, P)$ as $n \to \infty$.

To finish the proof we must show, with out assumptions on $C > -\infty$, that $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$ which will certainly imply $P(X_{-\infty} = \infty) = 0$. To prove this, notice that $X_{-\infty}^+ = \lim_{n\to-\infty} X_n^+$ and that by Jensen's inequality, $\{X_n^+\}_{n=1}^\infty$ is a non-negative backwards submartingale. Since $\inf \mathbb{E}X_n^+ \geq 0 > -\infty$, it follows by what we have just proved that $X_{-\infty}^+ \in L^1(\Omega, \mathcal{B}, P)$. ■

Remark 19.66. Let us give a direct proof of the fact that X is uniformly integrable if $C > -\infty$. We begin with Jensen's inequality;

$$\mathbb{E}|X_n| = 2\mathbb{E}X_n^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - \mathbb{E}X_n \leq 2\mathbb{E}X_0^+ - C = K < \infty, \quad (19.58)$$

which shows that $\{X_n\}_{n=1}^\infty$ is L^1 - bounded. For uniform integrability we will use the following identity;

$$\begin{aligned}\mathbb{E}[|X| : |X| \geq \lambda] &= \mathbb{E}[X : X \geq \lambda] - \mathbb{E}[X : X \leq -\lambda] \\ &= \mathbb{E}[X : X \geq \lambda] - (\mathbb{E}X - \mathbb{E}[X : X > -\lambda]) \\ &= \mathbb{E}[X : X \geq \lambda] + \mathbb{E}[X : X > -\lambda] - \mathbb{E}X.\end{aligned}$$

Taking $X = X_n$ and $k \geq n$, we find

$$\begin{aligned}\mathbb{E}[|X_n| : |X_n| \geq \lambda] &= \mathbb{E}[X_n : X_n \geq \lambda] + \mathbb{E}[X_n : X_n > -\lambda] - \mathbb{E}X_n \\ &\leq \mathbb{E}[X_k : X_n \geq \lambda] + \mathbb{E}[X_k : X_n > -\lambda] - \mathbb{E}X_k + (\mathbb{E}X_k - \mathbb{E}X_n) \\ &= \mathbb{E}[X_k : X_n \geq \lambda] - \mathbb{E}[X_k : X_n \leq -\lambda] + (\mathbb{E}X_k - \mathbb{E}X_n) \\ &= \mathbb{E}[|X_k| : |X_n| \geq \lambda] + (\mathbb{E}X_k - \mathbb{E}X_n).\end{aligned}$$

Given $\varepsilon > 0$ we may choose $k = k_\varepsilon < 0$ such that if $n \leq k$, $0 \leq \mathbb{E}X_k - \mathbb{E}X_n \leq \varepsilon$ and hence

$$\limsup_{\lambda\uparrow\infty} \sup_{n\leq k} \mathbb{E}[|X_n| : |X_n| \geq \lambda] \leq \limsup_{\lambda\uparrow\infty} \mathbb{E}[|X_k| : |X_n| \geq \lambda] + \varepsilon \leq \varepsilon$$

wherein we have used Eq. (19.58), Chebyschev's inequality to conclude $P(|X_n| \geq \lambda) \leq K/\lambda$ and then the uniform integrability of the singleton set, $\{|X_k|\} \subset L^1(\Omega, \mathcal{B}, P)$. From this it now easily follows that $\{X_n\}_{n\leq 0}$ is a uniformly integrable.

Corollary 19.67. *Suppose $1 \leq p < \infty$ and X_n in Theorem 19.65 is an L^p – bounded martingale. Then $X_n \to X_{-\infty}$ in $L^p(P)$. Moreover $X_{-\infty} = \mathbb{E}[X_0|\mathcal{B}_{-\infty}]$*

Proof. Since $X_n = \mathbb{E}[X_0|\mathcal{B}_n]$ for all n, it follows by cJensen that $|X_n|^p \leq \mathbb{E}[|X_0|^p|\mathcal{B}_n]$ for all n. By Proposition 19.7, $\{\mathbb{E}[|X_0|^p|\mathcal{B}_n]\}_{n\leq 0}$ is uniformly integrable and so is $\{|X_n|^p\}_{n\leq 0}$. By Theorem 19.65, $X_n \to X_{-\infty}$ a.s.. Hence we may now apply Corollary 11.34 to see that $X_n \to X_{-\infty}$ in $L^p(P)$. ■

Example 19.68 (SLLN). In this example we are going to give another proof of the strong law of large numbers in Theorem 12.44. Let $\{X_n\}_{n=1}^\infty$ be i.i.d. random variables such that $\mathbb{E}X_n = 0$ and let $S_{-n} := X_1 + \cdots + X_n$ and $\mathcal{B}_{-n} = \sigma(S_n, S_{n+1}, S_{n+2}, \dots)$ so that S_n is $\mathcal{B}_{-n}$ measurable for all n.

1. For any permutation σ of the set $\{1, 2, \ldots, n\}$,

$$(X_1, \ldots, X_n, S_n, S_{n+1}, S_{n+2}, \ldots) \overset{d}{=} (X_{\sigma 1}, \ldots, X_{\sigma n}, S_n, S_{n+1}, S_{n+2}, \ldots)$$

and in particular

$$(X_j, S_n, S_{n+1}, S_{n+2}, \ldots) \overset{d}{=} (X_1, S_n, S_{n+1}, S_{n+2}, \ldots) \text{ for all } j \le n.$$

2. By Exercise 18.5 we may conclude that

$$\mathbb{E}[X_j | S_n, S_{n+1}, S_{n+2}, \ldots] = \mathbb{E}[X_1 | S_n, S_{n+1}, S_{n+2}, \ldots] \text{ for all } j \le n. \tag{19.59}$$

3. Summing Eq. (19.59) over $j = 1, 2, \ldots, n$ gives,

$$S_n = \mathbb{E}[S_n | S_n, S_{n+1}, S_{n+2}, \ldots] = n\mathbb{E}[X_1 | S_n, S_{n+1}, S_{n+2}, \ldots]$$

from which it follows that

$$M_n := \frac{S_n}{n} := \mathbb{E}[X_1 | S_n, S_{n+1}, S_{n+2}, \ldots] \tag{19.60}$$

and hence $\left\{M_n = \frac{1}{n} S_n\right\}$ is a backwards martingale.

4. By Theorem 19.65we know;

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} M_n \text{ exists a.s.}$$

5. By Kolmogorov's zero one law (Proposition 7.41) we know that $\lim_{n\to\infty} \frac{S_n}{n} = c$ a.s. for some constant c.
6. Equation (19.60) along with Proposition 19.7 shows $\left\{\frac{S_n}{n}\right\}_{n=1}^{\infty}$ is uniformly integrable. Therefore,

$$\lim_{n\to\infty} \frac{S_n}{n} \overset{\text{a.s.}}{=} c = \mathbb{E}\left[\lim_{n\to\infty} \frac{S_n}{n}\right] = \lim_{n\to\infty} \mathbb{E}\left[\frac{S_n}{n}\right] = \mathbb{E}X_1$$

wherein we have use Theorem 11.31 to justify the interchange of the limit with the expectation. This shows $c = \mathbb{E}X_1$.

We have proved the strong law of large numbers.

19.9 Appendix: Some Alternate Proofs

This section may be safely omitted.

Proof. Alternate proof of Theorem 19.36. Let $A \in \mathcal{B}_\sigma$. Then

$$\mathbb{E}\left[X_\tau - X_\sigma : A\right] = \mathbb{E}\left[\sum_{k=0}^{N-1} 1_{\sigma \le k < \tau} \Delta_{k+1} X : A\right]$$
$$= \sum_{k=1}^{N} \mathbb{E}\left[\Delta_k X : A \cap \{\sigma \le k < \tau\}\right].$$

Since $A \in \mathcal{B}_\sigma$, $A \cap \{\sigma \le k\} \in \mathcal{B}_k$ and since $\{k < \tau\} = \{\tau \le k\}^c \in \mathcal{B}_k$, it follows that $A \cap \{\sigma \le k < \tau\} \in \mathcal{B}_k$. Hence we know that

$$\mathbb{E}\left[\Delta_{k+1} X : A \cap \{\sigma \le k < \tau\}\right] \overset{\le}{\underset{\ge}{=}} 0 \text{ respectively.}$$

and hence that

$$\mathbb{E}\left[X_\tau - X_\sigma : A\right] \overset{\le}{\underset{\ge}{=}} 0 \text{ respectively.}$$

Since this true for all $A \in \mathcal{B}_\sigma$, Eq. (19.21) follows. ∎

Lemma 19.69. *Suppose $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ is a filtered probability space, $1 \le p < \infty$, and let $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma\left(\cup_{n=1}^\infty \mathcal{B}_n\right)$. Then $\cup_{n=1}^\infty L^p(\Omega, \mathcal{B}_n, P)$ is dense in $L^p(\Omega, \mathcal{B}_\infty, P)$.*

Proof. Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$, then M_n is an increasing sequence of closed subspaces of $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. Further let $\mathbb{A}$ be the algebra of functions consisting of those $f \in \cup_{n=1}^\infty M_n$ such that f is bounded. As a consequence of the density Theorem 9.8, we know that $\mathbb{A}$ and hence $\cup_{n=1}^\infty M_n$ is dense in $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. This completes the proof. However for the readers convenience let us quickly review the proof of Theorem 9.8 in this context.

Let $\mathbb{H}$ denote those bounded $\mathcal{B}_\infty$ – measurable functions, $f : \Omega \to \mathbb{R}$, for which there exists $\{\varphi_n\}_{n=1}^\infty \subset \mathbb{A}$ such that $\lim_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} = 0$. A routine check shows $\mathbb{H}$ is a subspace of the bounded $\mathcal{B}_\infty$ –measurable $\mathbb{R}$ – valued functions on Ω, $1 \in \mathbb{H}$, $\mathbb{A} \subset \mathbb{H}$ 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 (or bounded) convergence theorem, $\lim_{n\to\infty} \|(f - f_n)\|_{L^p(P)} = 0$.[7] We may now choose $\varphi_n \in \mathbb{A}$ such that $\|\varphi_n - f_n\|_{L^p(P)} \le \frac{1}{n}$ then

$$\limsup_{n\to\infty} \|f - \varphi_n\|_{L^p(P)} \le \limsup_{n\to\infty} \|(f - f_n)\|_{L^p(P)} + \limsup_{n\to\infty} \|f_n - \varphi_n\|_{L^p(P)} = 0,$$

which implies $f \in \mathbb{H}$.

An application of Dynkin's Multiplicative System Theorem 9.3, now shows $\mathbb{H}$ contains all bounded $\sigma(\mathbb{A}) = \mathcal{B}_\infty$ – measurable functions on Ω. Since for any $f \in L^p(\Omega, \mathcal{B}, P)$, $f 1_{|f| \le n} \in \mathbb{H}$ there exists $\varphi_n \in \mathbb{A}$ such that $\|f_n - \varphi_n\|_p \le n^{-1}$. Using the DCT we know that $f_n \to f$ in L^p and therefore by Minikowski's inequality it follows that $\varphi_n \to f$ in L^p. ∎

[7] It is at this point that the proof would break down if $p = \infty$.

Theorem 19.70. *Suppose* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ *is a filtered probability space,* $1 \le p < \infty$, *and let* $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$. *Then for every* $X \in L^p(\Omega, \mathcal{B}, P)$, $X_n = \mathbb{E}[X|\mathcal{B}_n]$ *is a martingale and* $X_n \to X_\infty := \mathbb{E}[X|\mathcal{B}_\infty]$ *in* $L^p(\Omega, \mathcal{B}_\infty, P)$ *as* $n \to \infty$.

Proof. We have already seen in Example 19.5 that $X_n = \mathbb{E}[X|\mathcal{B}_n]$ is always a martingale. Since conditional expectation is a contraction on L^p it follows that $\mathbb{E}|X_n|^p \le \mathbb{E}|X|^p < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$. So to finish the proof we need to show $X_n \to X_\infty$ in $L^p(\Omega, \mathcal{B}, P)$ as $n \to \infty$.

Let $M_n := L^p(\Omega, \mathcal{B}_n, P)$ and $M_\infty = L^p(\Omega, \mathcal{B}_\infty, P)$. If $X \in \cup_{n=1}^\infty M_n$, then $X_n = X$ for all sufficiently large n and for $n = \infty$. Now suppose that $X \in M_\infty$ and $Y \in \cup_{n=1}^\infty M_n$. Then

$$\begin{aligned}\|\mathbb{E}_{\mathcal{B}_\infty} X - \mathbb{E}_{\mathcal{B}_n} X\|_p &\le \|\mathbb{E}_{\mathcal{B}_\infty} X - \mathbb{E}_{\mathcal{B}_\infty} Y\|_p + \|\mathbb{E}_{\mathcal{B}_\infty} Y - \mathbb{E}_{\mathcal{B}_n} Y\|_p + \|\mathbb{E}_{\mathcal{B}_n} Y - \mathbb{E}_{\mathcal{B}_n} X\|_p \\ &\le 2\|X - Y\|_p + \|\mathbb{E}_{\mathcal{B}_\infty} Y - \mathbb{E}_{\mathcal{B}_n} Y\|_p\end{aligned}$$

and hence

$$\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{B}_\infty} X - \mathbb{E}_{\mathcal{B}_n} X\|_p \le 2\|X - Y\|_p.$$

Using the density Lemma 19.69 we may choose $Y \in \cup_{n=1}^\infty M_n$ as close to $X \in M_\infty$ as we please and therefore it follows that $\limsup_{n\to\infty} \|\mathbb{E}_{\mathcal{B}_\infty} X - \mathbb{E}_{\mathcal{B}_n} X\|_p = 0$.

For general $X \in L^p(\Omega, \mathcal{B}, P)$ it suffices to observe that $X_\infty := \mathbb{E}[X|\mathcal{B}_\infty] \in L^p(\Omega, \mathcal{B}_\infty, P)$ and by the tower property of conditional expectations,

$$\mathbb{E}[X_\infty|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n] = \mathbb{E}[X|\mathcal{B}_n] = X_n.$$

So again $X_n \to X_\infty$ in L^p as desired. ■

We are now ready to prove the converse of Theorem 19.70.

Theorem 19.71. *Suppose* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^\infty, P)$ *is a filtered probability space,* $1 \le p < \infty$, $\mathcal{B}_\infty := \vee_{n=1}^\infty \mathcal{B}_n := \sigma(\cup_{n=1}^\infty \mathcal{B}_n)$, *and* $\{X_n\}_{n=1}^\infty \subset L^p(\Omega, \mathcal{B}, P)$ *is a martingale. Further assume that* $\sup_n \|X_n\|_p < \infty$ *and that* $\{X_n\}_{n=1}^\infty$ *is uniformly integrable if* $p = 1$. *Then there exists* $X_\infty \in L^p(\Omega, \mathcal{B}_\infty, P)$ *such that* $X_n := \mathbb{E}[X_\infty|\mathcal{B}_\infty]$. *Moreover by Theorem 19.70 we know that* $X_n \to X_\infty$ *in* $L^p(\Omega, \mathcal{B}_\infty, P)$ *as* $n \to \infty$ *and hence* X_∞ *is uniquely determined by* $\{X_n\}_{n=1}^\infty$.

Proof. By Theorems 16.19 and 16.21, there exists $X_\infty \in L^p(\Omega, \mathcal{B}_\infty, P)$ and a subsequence, $Y_k = X_{n_k}$ such that

$$\lim_{k\to\infty} \mathbb{E}[Y_k h] = \mathbb{E}[X_\infty h] \text{ for all } h \in L^q(\Omega, \mathcal{B}_\infty, P)$$

where $q := p(p-1)^{-1}$. Using the martingale property, if $h \in (\mathcal{B}_n)_b$ for some n, it follows that $\mathbb{E}[Y_k h] = \mathbb{E}[X_n h]$ for all large k and therefore that

$$\mathbb{E}[X_\infty h] = \mathbb{E}[X_n h] \text{ for all } h \in (\mathcal{B}_n)_b.$$

This implies that $X_n = \mathbb{E}[X_\infty|\mathcal{B}_n]$ as desired. ■

Theorem 19.72 (Almost sure convergence). *Suppose* $(\Omega, \mathcal{B}, \{\mathcal{B}_n\}_{n=0}^{\infty}, P)$ *is a filtered probability space,* $1 \leq p < \infty$, *and let* $\mathcal{B}_\infty := \vee_{n=1}^{\infty} \mathcal{B}_n := \sigma(\cup_{n=1}^{\infty} \mathcal{B}_n)$. *Then for every* $X \in L^1(\Omega, \mathcal{B}, P)$, *the martingale,* $X_n = \mathbb{E}[X|\mathcal{B}_n]$, *converges almost surely to* $X_\infty := \mathbb{E}[X|\mathcal{B}_\infty]$.

Before starting the proof, recall from Proposition 1.5, if $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are two bounded sequences, then

$$\begin{aligned}\limsup_{n\to\infty}(a_n+b_n) - \liminf_{n\to\infty}(a_n+b_n)& \\ \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n &- \left(\liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n\right) \\ = \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} a_n &+ \limsup_{n\to\infty} b_n - \liminf_{n\to\infty} b_n. \quad (19.61)\end{aligned}$$

Proof. Since

$$X_n = \mathbb{E}[X|\mathcal{B}_n] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_\infty]|\mathcal{B}_n] = \mathbb{E}[X_\infty|\mathcal{B}_n],$$

there is no loss in generality in assuming $X = X_\infty$. If $X \in M_n := L^1(\Omega, \mathcal{B}_n, P)$, then $X_m = X_\infty$ a.s. for all $m \geq n$ and hence $X_m \to X_\infty$ a.s. Therefore the theorem is valid for any X in the dense (by Lemma 19.69) of $L^1(\Omega, \mathcal{B}_\infty, P)$.

For general $X \in L^1(\Omega, \mathcal{B}_\infty, P)$, let $Y_j \in \cup M_n$ such that $Y_j \to X \in L^1(\Omega, \mathcal{B}_\infty, P)$ and let $Y_{j,n} := \mathbb{E}[Y_j|\mathcal{B}_n]$ and $X_n := \mathbb{E}[X|\mathcal{B}_n]$. We know that $Y_{j,n} \to Y_{j,\infty}$ a.s. for each $j \in \mathbb{N}$ and our goal is to show $X_n \to X_\infty$ a.s. By Doob's inequality in Corollary 19.43 and the L^1 - contraction property of conditional expectation we know that

$$P(X_N^* \geq a) \leq \frac{1}{a}\mathbb{E}|X_N| \leq \frac{1}{a}\mathbb{E}|X|$$

and so passing to the limit as $N \to \infty$ we learn that

$$P\left(\sup_n |X_n| \geq a\right) \leq \frac{1}{a}\mathbb{E}|X| \text{ for all } a > 0. \quad (19.62)$$

Letting $a \uparrow \infty$ then shows $P(\sup_n |X_n| = \infty) = 0$ and hence $\sup_n |X_n| < \infty$ a.s. Hence we may use Eq. (19.61) with $a_n = X_n - Y_{j,n}$ and $b_n := Y_{j,n}$ to find

$$\begin{aligned}D &= \limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n \\ &\leq \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n - \liminf_{n\to\infty} b_n \\ &= \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} a_n \leq 2\sup_n |a_n| \\ &= 2\sup_n |X_n - Y_{j,n}|,\end{aligned}$$

wherein we have used $\limsup_{n\to\infty} b_n - \liminf_{n\to\infty} b_n = 0$ a.s. since $Y_{j,n} \to Y_{j,\infty}$ a.s.

We now apply Doob's inequality one more time, i.e. use Eq. (19.62) with $X_n \to X_n - Y_{j,n}$ and $X \to X - Y_j$, to conclude,

$$P(D \geq a) \leq P\left(\sup_n |X_n - Y_{j,n}| \geq \frac{a}{2}\right) \leq \frac{2}{a}\mathbb{E}|X - Y_j| \to 0 \text{ as } j \to \infty.$$

Since $a > 0$ is arbitrary here, it follows that $D = 0$ a.s., i.e. $\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n$ and hence $\lim_{n\to\infty} X_n$ exists in $\mathbb{R}$ almost surely. Since we already know that $X_n \to X_\infty$ in $L^1(\Omega, \mathcal{B}, P)$, we may conclude that $\lim_{n\to\infty} X_n = X_\infty$ a.s. ∎