Bruce K. Driver

Math 280 (Probability Theory) Lecture Notes

November 2, 2006 *File:prob.tex*

Contents

Part I

Background Material

Limsups, Liminfs and Extended Limits

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

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

1. *If* $a_n \leq b_n$ *for*[1] *a.a.* n *then* $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
2. *If* $c \in \mathbb{R}$, $\lim_{n\to\infty}(ca_n) = c\lim_{n\to\infty} a_n$.
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 $a_n = n$ 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* $\limsup_{n\to\infty}$.

[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 $n \geq N$.

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

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

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

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

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

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

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

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

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

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

i.e.

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

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

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

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

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

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

From this we learn that

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

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

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

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

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

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

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

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

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

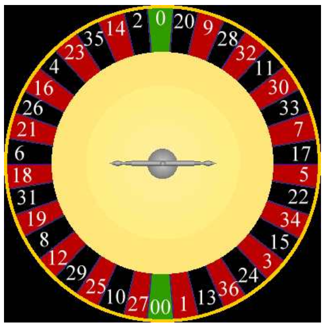
$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{kn} = M.$$

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, $0 \le t < \infty$. In this case, we might take,
$$\Omega = \left\{\omega \in C\left([0,\infty), \mathbb{R}^3\right) : \omega(0) = x\right\}.$$

Definition 2.3. *An event 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. $A = \cap_{N=1}^{\infty} \cup_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where there are infinitely many heads tossed in the sequence.
4. $A = \cup_{N=1}^{\infty} \cap_{n \ge N} \{\omega \in \Omega : \omega_n = 1\}$ is the event where heads occurs from some time onwards, i.e. $\omega \in A$ iff there exists, $N = N(\omega)$ such that $\omega_n = 1$ for all $n \ge N$.

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

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

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:

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

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

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

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

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

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

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

we see that

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

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

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

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

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

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

Proof. We are going to use the fact proved below in 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

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

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

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

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

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

and

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

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

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

and this may be expressed as

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

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

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

which may be written as

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

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

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

be the 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. $\sup_{k\geq n} 1_{A_k}(x) = 1_{\cup_{k\geq n} A_k} = 1_{\sup_{k\geq n} A_n}$,
5. $\inf 1_{A_k}(x) = 1_{\cap_{k\geq n} A_k} = 1_{\inf_{k\geq n} A_k}$,
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 $A_1 \cup \dots \cup A_n \in \mathcal{A}$. In view of conditions 1. and 2., 3. is equivalent to*

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

Definition 3.14. *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

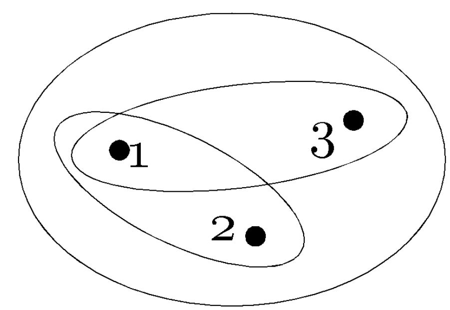


Fig. 3.1. A collection of subsets.

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

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 $\mathcal{E} = \{A_1, \dots, A_n\}$ where $A_1, \dots, A_n$ is a partition of X. In this case

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

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

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

Example 3.20. 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

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

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 σ – field on $\mathbb{R}$. Then

$$\mathcal{E}_{(0,1]} = \{(a,b] : 0 \le a < b \le 1\}$$

and we have

$$\mathcal{B}_{(0,1]} = \sigma\left(\mathcal{E}_{(0,1]}\right).$$

In particular, if $A \in \mathcal{B}$ such that $A \subset (0,1]$, then $A \in \sigma\left(\mathcal{E}_{(0,1]}\right)$.

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

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**) $\mu(E) \le \mu(F)$ 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**) $\mu(\cup_{j=1}^n E_j) \le \sum_{j=1}^n \mu(E_j)$.*
4. *μ is sub-additive on $\mathcal{A}$ iff*
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A = \sum_{i=1}^{\infty} A_i \tag{4.4}$$
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*
$$\mu\left(\sum_{i=1}^{\infty} A_i\right) \ge \sum_{i=1}^{\infty} \mu(A_i).$$
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
$$\mu(F) = \mu(E) + \mu(F \setminus E) \ge \mu(E).$$
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 $\widetilde{E}_j = E_j \setminus (E_1 \cup \dots \cup E_{j-1})$ 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
$$\mu(E) = \sum \mu(\widetilde{E}_j) \le \sum \mu(E_j).$$

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 $A_i := B_i \setminus (B_1 \cup \dots B_{i-1}) \in \mathcal{A}$ and $B_0 = \emptyset$. Therefore using the monotonicity of μ and Eq. (4.4)
$$\mu(A) \le \sum_{i=1}^{\infty} \mu(A_i) \le \sum_{i=1}^{\infty} \mu(B_i).$$
5. Suppose that $A = \sum_{i=1}^{\infty} A_i$ with $A_i, A \in \mathcal{A}$, then $\sum_{i=1}^{n} A_i \subset A$ for all n and so by the monotonicity and finite additivity of μ, $\sum_{i=1}^{n} \mu(A_i) \le \mu(A)$. Letting $n \to \infty$ in this equation shows μ is superadditive.
6. This is a combination of items 5. and 6.

■

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

$2 \implies 1$. 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 $n \ge 2$, let $A_n' := A_n \setminus (A_1 \cup \dots A_{n-1}) \in \mathcal{B}$. Then
$$P(\cup_{n=1}^{\infty} A_n) = P(\cup_{n=1}^{\infty} A_n') = \sum_{n=1}^{\infty} P(A_n') \le \sum_{n=1}^{\infty} P(A_n).$$

■

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 σ − 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

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

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
$$\begin{aligned} F(-\infty) &:= \lim_{x \to -\infty} F(x) = 0 \text{ and} \\ F(+\infty) &:= \lim_{x \to \infty} F(x) = 1 \end{aligned}$$
there exists a finitely additive probability measure, $P = P_F$ on $\mathcal{A}$ such that
$$P((a, b] \cap \mathbb{R}) = F(b) - F(a) \text{ for all } -\infty \le a \le b \le \infty.$$
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 $F(a+) := \lim_{x \downarrow a} F(x) \ne F(a)$, 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*

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

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

$$\begin{aligned}\mathbb{E}(f)&=\sum_{y\in\mathbb{C}}yP(f=y)=\lim_{N\to\infty}\frac{1}{N}\sum_{y\in\mathbb{C}}y\cdot\#\{1\le k\le N:f(\omega(k))=y\}\\&=\lim_{N\to\infty}\frac{1}{N}\sum_{y\in\mathbb{C}}y\cdot\sum_{k=1}^{N}1_{f(\omega(k))=y}=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\sum_{y\in\mathbb{C}}f(\omega(k))\cdot1_{f(\omega(k))=y}\\&=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}f(\omega(k)).\end{aligned}$$

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. $\mathbb{E}(f)\ge0$ 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 $\lambda\neq0$, 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
$$\begin{aligned}\mathbb{E}(f+g)&=\sum_{z\in\mathbb{C}}z\ P(f+g=z)\\&=\sum_{z\in\mathbb{C}}z\ P\left(\cup_{a+b=z}\{f=a,\ g=b\}\right)\\&=\sum_{z\in\mathbb{C}}z\sum_{a+b=z}P(\{f=a,\ g=b\})\\&=\sum_{z\in\mathbb{C}}\sum_{a+b=z}(a+b)P(\{f=a,\ g=b\})\\&=\sum_{a,b}(a+b)P(\{f=a,\ g=b\}).\end{aligned}$$
But
$$\begin{aligned}\sum_{a,b}aP(\{f=a,\ g=b\})&=\sum_a a\sum_b P(\{f=a,\ g=b\})\\&=\sum_a aP\left(\cup_b\{f=a,\ g=b\}\right)\\&=\sum_a aP(\{f=a\})=\mathbb{E}f\end{aligned}$$
and similarly,
$$\sum_{a,b}bP(\{f=a,\ g=b\})=\mathbb{E}g.$$
Equation (4.9) is now a consequence of the last three displayed equations.
3. If $f\ge0$ then
$$\mathbb{E}(f)=\sum_{a\ge0}aP(f=a)\ge0.$$

4. First observe that
$$|f| = \sum_{\lambda\in\mathbb{C}} |\lambda| \, 1_{f=\lambda}$$
and therefore,
$$\mathbb{E}|f| = \mathbb{E}\sum_{\lambda\in\mathbb{C}} |\lambda| \, 1_{f=\lambda} = \sum_{\lambda\in\mathbb{C}} |\lambda| \, \mathbb{E}1_{f=\lambda} = \sum_{\lambda\in\mathbb{C}} |\lambda| \, P(f=\lambda) \le \max|f|.$$
On the other hand,
$$|\mathbb{E}f| = \left|\sum_{\lambda\in\mathbb{C}} \lambda P(f=\lambda)\right| \le \sum_{\lambda\in\mathbb{C}} |\lambda| \, P(f=\lambda) = \mathbb{E}|f|.$$
■

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
$$P(\{|f| \ge \varepsilon\}) = \mathbb{E}\left[1_{|f|\ge\varepsilon}\right] \le \mathbb{E}\left[\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}\right] \le \varepsilon^{-p}\mathbb{E}|f|^p. \tag{4.11}$$
Observe that
$$|f|^p = \sum_{\lambda\in\mathbb{C}} |\lambda|^p \, 1_{\{f=\lambda\}}$$
is a simple random variable and $\{|f| \ge \varepsilon\} = \sum_{|\lambda|\ge\varepsilon} \{f=\lambda\} \in \mathcal{A}$ as well. Therefore, $\frac{|f|^p}{\varepsilon^p} 1_{|f|\ge\varepsilon}$ is still a simple random variable.

Lemma 4.19 (Inclusion Exclusion Formula). *If $A_n \in \mathcal{A}$ for $n = 1, 2, \dots, M$ such that $\mu\left(\cup_{n=1}^M A_n\right) < \infty$, then*
$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le M} \mu\left(A_{n_1} \cap \dots \cap A_{n_k}\right). \tag{4.12}$$

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
$$\begin{aligned} 1 - 1_A = 1_{A^c} &= \prod_{n=1}^M 1_{A_n^c} = \prod_{n=1}^M (1 - 1_{A_n}) \\ &= \sum_{k=0}^M (-1)^k \sum_{0\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}} \cdots 1_{A_{n_k}} \\ &= \sum_{k=0}^M (-1)^k \sum_{0\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}\cap\dots\cap A_{n_k}} \end{aligned}$$
from which it follows that
$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}\cap\dots\cap A_{n_k}}. \tag{4.13}$$
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 $\omega \in A_1 \cap \dots \cap A_m$ and $\omega \notin A_{m+1} \cup \dots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (4.13) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have
$$\begin{aligned} &\sum_{k=1}^M (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le M} 1_{A_{n_1}\cap\dots\cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \sum_{1\le n_1<n_2<\dots<n_k\le m} 1_{A_{n_1}\cap\dots\cap A_{n_k}}(\omega) \\ &= \sum_{k=1}^m (-1)^{k+1} \binom{m}{k} \\ &= 1 - \sum_{k=0}^m (-1)^k (1)^{n-k} \binom{m}{k} \\ &= 1 - (1-1)^m = 1. \end{aligned}$$
This verifies Eq. (4.13) 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), $\omega : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$, and define $P(A) := \frac{\#(A)}{n!}$ to be the uniform distribution (Haar measure) on Ω. We wish to compute the probability of the event, B, that a random permutation fixes some index i. To do this, let $A_i := \{\omega \in \Omega : \omega(i) = i\}$ and observe that $B = \cup_{i=1}^n A_i$. So by the Inclusion Exclusion Formula, we have

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

Since

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

and

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

we find

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

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 $i = 1, 2, \dots, n$. We further suppose $p : \Omega \to [0,1]$ is a function such that

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

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

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

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 $i = 1, 2, \dots, n$ 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
$$P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{4.18}$$

   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,

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

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

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

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

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

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

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

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

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

Hence we find

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

where

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

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

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

and hence we may conclude that

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

and therefore, that

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

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

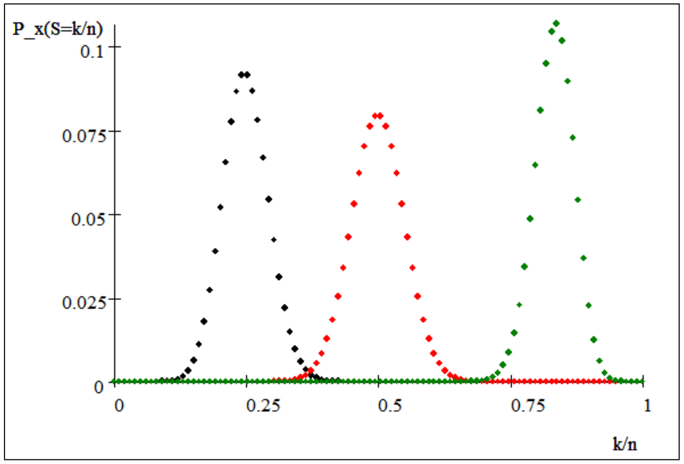


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

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 $i = 1, 2, \dots, n$, Λ_i is a finite collection of disjoint sets from $\mathcal{S}$. Then

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

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

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

■

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,

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

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*

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

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,

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

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,

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

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

$$\mu(A) = \sum_{j=1}^{k} \mu(E_j) \le \sum_{j=1}^{k} \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu\left(E_{n,i} \cap E_j\right)$$
$$= \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \sum_{j=1}^{k} \mu\left(E_{n,i} \cap E_j\right) = \sum_{n=1}^{\infty} \sum_{i=1}^{N_n} \mu\left(E_{n,i}\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right),$$

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

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

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,

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

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

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

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

[1] To see this, let $c := \sup\left\{x \le b : [\tilde{a}, x] \text{ is finitely covered by } \left\{\tilde{J}_n^o\right\}_{n=1}^{\infty}\right\}$. 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.

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

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,

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

Since $C \setminus A^N \downarrow C \setminus A$, it also follows that $\mu\left(C \setminus A^N\right) < \varepsilon$ for sufficiently large N and this shows $B = \cup_{i=1}^\infty B_i \in \mathcal{B}_0$. Hence $\mathcal{B}_0$ is a sub-σ-algebra of $\mathcal{B} = \sigma(\mathcal{A})$ which contains $\mathcal{A}$ which shows $\mathcal{B}_0 = \mathcal{B}$. ■

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a $\sigma-$ 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

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

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

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

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 $\mu(C_n \setminus B) \le 1/n$ and $\mu(B \setminus A_n) \le 1/n$. By replacing A_N by $\cup_{n=1}^N A_n$ and C_N by $\cap_{n=1}^N C_n$, we may assume that $A_n \uparrow$ and $C_n \downarrow$ as n increases. Let $A = \cup A_n \in \mathcal{A}_{\delta\sigma}$ and $C = \cap C_n \in \mathcal{A}_{\sigma\delta}$, then $A \subset B \subset C$ and

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

■

Exercise 5.2. Let $\mathcal{B} = \mathcal{B}_{\mathbb{R}^n} = \sigma(\{\text{open subsets of } \mathbb{R}^n\})$ be the Borel σ – algebra on $\mathbb{R}^n$ and μ be a probability measure on $\mathcal{B}$. Further, let $\mathcal{B}_0$ denote those sets $B \in \mathcal{B}$ such that for every $\varepsilon > 0$ there exists $F \subset B \subset V$ such that F is closed, V is open, and $\mu(V \setminus F) < \varepsilon$. Show:

1. $\mathcal{B}_0$ contains all closed subsets of $\mathcal{B}$. **Hint:** given a closed subset, $F \subset \mathbb{R}^n$ and $k \in \mathbb{N}$, let $V_k := \cup_{x \in F} B(x, 1/k)$, where $B(x,\delta) := \{y \in \mathbb{R}^n : |y - x| < \delta\}$. Show, $V_k \downarrow F$ as $k \to \infty$.
2. Show $\mathcal{B}_0$ is a σ – algebra and use this along with the first part of this exercise to conclude $\mathcal{B} = \mathcal{B}_0$. **Hint:** follow closely the method used in the first step of the proof of Theorem 5.10.
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 $\mu(B \setminus K) < \varepsilon$. **Hint:** take $K := F \cap \{x \in \mathbb{R}^n : |x| \le n\}$ 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 $\mu(A_n) \uparrow \mu(A)$ for all $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ such that $A_n \uparrow A \in \mathcal{A}$. Furthermore if $\mu(X) < \infty$, 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 $\mu(A_n) \uparrow \mu(A)$ as $n \to \infty$.*
2. ***(Monotonicity)*** *If $A, B \in \mathcal{A}_\sigma$ with $A \subset B$ then $\mu(A) \le \mu(B)$.*
3. ***(Strong Additivity)*** *If $A, B \in \mathcal{A}_\sigma$, then*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{5.3}$$
4. ***(Sub-Additivity on $\mathcal{A}_\sigma$)*** *The function μ is sub-additive on $\mathcal{A}_\sigma$, i.e. if $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$, then*
$$\mu(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.4}$$
5. *(σ - Additivity **on $\mathcal{A}_\sigma$**) The function μ is countably additive on $\mathcal{A}_\sigma$.*

Proof. 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
$$\begin{aligned}\mu(A) &= \mu(X) - \mu(A^c) = \mu(X) - \lim_{n\to\infty} \mu(A_n^c) \\ &= \lim_{n\to\infty} \left[\mu(X) - \mu(A_n^c)\right] = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$
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,
$$\begin{aligned}\mu(A \cup B) &= \lim_{n\to\infty} \mu(A_n \cup B_n) = \lim_{n\to\infty} \left[\mu(A_n) + \mu(B_n) - \mu(A_n \cap B_n)\right] \\ &= \mu(A) + \mu(B)\end{aligned}$$
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
$$\begin{aligned}\mu(X) + \mu(C \setminus A) &= \mu(A^c \cup C) + \mu(A^c \cap C) \\ &= \mu(A^c) + \mu(C) = \mu(X) - \mu(A) + \mu(C).\end{aligned}$$

■

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

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

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

Because $\mathcal{B}$ is an algebra, to verify that $\mathcal{B}$ is a σ – algebra it suffices to show that $B = \sum_{n=1}^\infty B_n \in \mathcal{B}$ whenever $\{B_n\}_{n=1}^\infty$ is a disjoint sequence in $\mathcal{B}$. To prove $B \in \mathcal{B}$, let $\varepsilon > 0$ be given and choose $A_i \subset B_i \subset C_i$ such that $A_i \in \mathcal{A}_\delta$, $C_i \in \mathcal{A}_\sigma$, and $\mu(C_i \setminus A_i) < \varepsilon 2^{-i}$ for all i. Since the $\{A_i\}_{i=1}^\infty$ are pairwise disjoint we may use Lemma 5.16 to show,

$$\sum_{i=1}^{n}\mu(C_i)=\sum_{i=1}^{n}(\mu(A_i)+\mu(C_i\setminus A_i))$$
$$=\mu(\cup_{i=1}^{n}A_i)+\sum_{i=1}^{n}\mu(C_i\setminus A_i)\le\mu(X)+\sum_{i=1}^{n}\varepsilon 2^{-i}.$$

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

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

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),

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

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

$$|\bar\mu(B)-\mu(A^n)|<\varepsilon+\sum_{i=n+1}^{\infty}\mu(C_i)\text{ and}$$
$$\left|\sum_{i=1}^{n}\bar\mu(B_i)-\mu(A^n)\right|=\left|\sum_{i=1}^{n}[\bar\mu(B_i)-\mu(A_i)]\right|\le\sum_{i=1}^{n}|\bar\mu(B_i)-\mu(A_i)|\le\varepsilon\sum_{i=1}^{n}2^{-i}<\varepsilon.$$

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,

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

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

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

Thus

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

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

5.5 Completions of Measure Spaces

Definition 5.21. *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 $\mu(f\ne g)=0$.

Definition 5.22. *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 5.23 (Completion of a Measure). *Let $(X, \mathcal{B}, \mu)$ be a measure space. Set*

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

*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

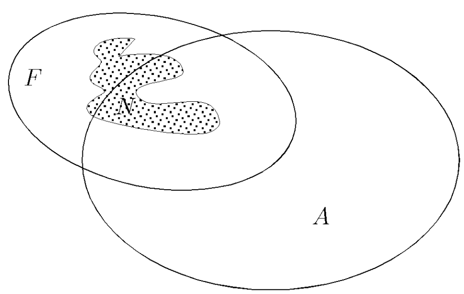


Fig. 5.1. Completing a σ – algebra.

that $N \subset F$ and $\mu(F) = 0$. Since $N^c = (F \setminus N) \cup F^c$,

$$(A \cup N)^c = A^c \cap N^c = A^c \cap (F \setminus N \cup F^c)$$
$$= [A^c \cap (F \setminus N)] \cup [A^c \cap F^c]$$

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 $\mu(\cup F_i) \le \sum \mu(F_i) = 0$. Therefore, $\bar{\mathcal{B}}$ is a σ – algebra. Suppose $A \cup N_1 = B \cup N_2$ with $A, B \in \mathcal{B}$ and $N_1, N_2, \in \mathcal{N}$. Then $A \subset A \cup N_1 \subset A \cup N_1 \cup F_2 = B \cup F_2$ which shows that

$$\mu(A) \le \mu(B) + \mu(F_2) = \mu(B).$$

Similarly, we show that $\mu(B) \le \mu(A)$ 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 $k \ge 0$.

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

Lemma 5.24. *Suppose $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \ne \emptyset$.*

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 $k \le n$, it follows that

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } k \le n. \tag{5.12}$$

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 $n \ge k$, we learn from Eq. (5.12) 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 5.25 (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 5.24, 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 $(\lambda_1, \ldots, \lambda_n) \in \Lambda^n$, let

$$p_n(\lambda_1, \ldots, \lambda_n) := P(\{\omega \in \Omega : \omega_1 = \lambda_1, \ldots, \omega_n = \lambda_n\}). \tag{5.13}$$

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 $(\lambda_1, \ldots, \lambda_n) \in \Lambda^n$,

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

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. (5.13) holds for all $n \in \mathbb{N}$ and $(\lambda_1, \ldots, \lambda_n) \in \Lambda^n$.

Example 5.26 (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 $(\lambda_1, \ldots, \lambda_n) \in \Lambda^n$, we have

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

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

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

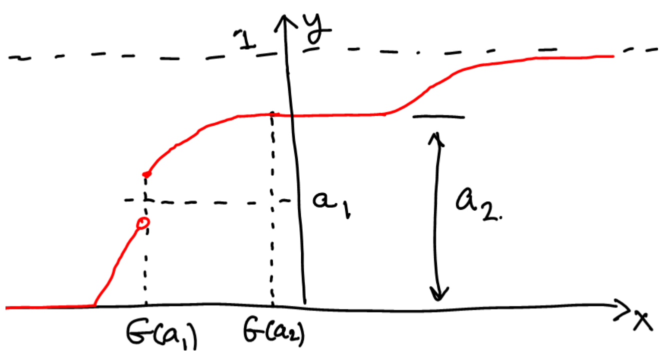


Fig. 6.1. A pictorial definition of G.

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.

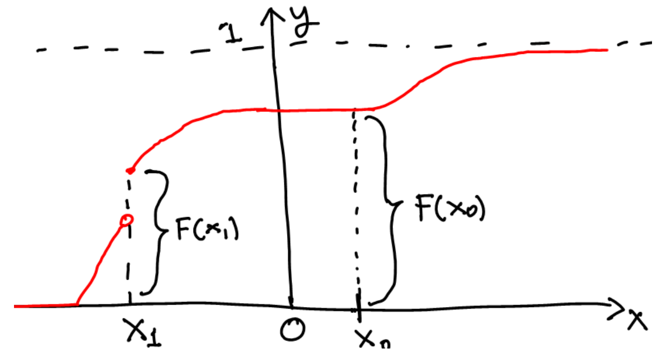


Fig. 6.2. As can be seen from this picture, $G(y) \leq x_0$ iff $y \leq F(x_0)$ and similary, $G(y) \leq x_1$ iff $y \leq x_1$.

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

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

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. ■

Lemma 6.12 (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 6.13 (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 6.14. Suppose that Y is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1, \ldots, y_N) = y_i$. Then, as the reader should check,

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

Proposition 6.15. *Assuming the notation in Definition 6.13 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 6.12. ($\Leftarrow$) 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 6.16. *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 6.17. *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 6.18. For $i = 1, 2, \ldots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each π_i is continuous and therefore $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable.

Lemma 6.19. *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \ldots, \pi_n)$ and in particular, $A_1 \times \cdots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \ldots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$.*

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

$$(a, b) = (a_1, b_1) \times \cdots \times (a_n, b_n). \tag{6.2}$$

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. (6.2), 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 6.20. *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 6.19 and Proposition 6.15. ■

Corollary 6.21. *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. ■

As an example of this material, let us give another proof of the existence of iid simple random variables – see Example 5.26 above.

Theorem 6.22 (Existence of i.i.d simple R.V.'s). *This Theorem has been moved to Theorem 7.22 below.*

Corollary 6.23 (Independent variables on product spaces). *This Corollary has been moved to Corollary 7.23 below.*

Lemma 6.24. *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

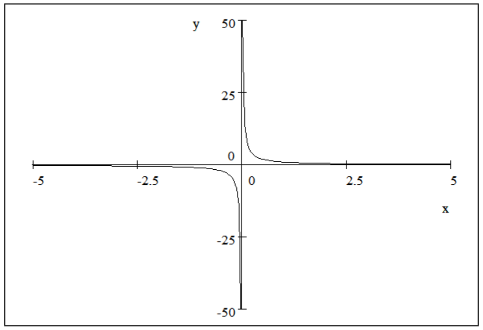
$$i(z) = \begin{cases} \frac{1}{z} \text{ if } & z \neq 0 \\ 0 \text{ if } & z = 0. \end{cases}$$

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 6.25. For the real case of Lemma 6.24, define i as above but now take z to real. From the plot of i, Figure 6.25, 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

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma(\{[a, \infty] : a \in \mathbb{R}\}). \tag{6.3}$$

Proposition 6.26 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{6.4}$$

In particular $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ and $\mathcal{B}_{\mathbb{R}} \subset \mathcal{B}_{\bar{\mathbb{R}}}$.

Proof. Let us first observe that

$$\{-\infty\} = \cap_{n=1}^{\infty} [-\infty, -n) = \cap_{n=1}^{\infty} [-n, \infty]^c \in \mathcal{B}_{\bar{\mathbb{R}}},$$
$$\{\infty\} = \cap_{n=1}^{\infty} [n, \infty] \in \mathcal{B}_{\bar{\mathbb{R}}} \text{ and } \mathbb{R} = \bar{\mathbb{R}} \setminus \{\pm\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}.$$

Letting $i : \mathbb{R} \to \bar{\mathbb{R}}$ be the inclusion map,

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

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. $A \in \mathcal{B}_{\bar{\mathbb{R}}} \Longrightarrow A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and
2. if $A \subset \bar{\mathbb{R}}$ is such that $A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ there exists $B \in \mathcal{B}_{\bar{\mathbb{R}}}$ such that $A \cap \mathbb{R} = B \cap \mathbb{R}$. Because $A \Delta B \subset \{\pm\infty\}$ and $\{\infty\}, \{-\infty\} \in \mathcal{B}_{\bar{\mathbb{R}}}$ we may conclude that $A \in \mathcal{B}_{\bar{\mathbb{R}}}$ as well.

This proves Eq. (6.4). ■

The proofs of the next two corollaries are left to the reader, see Exercises 6.5 and 6.6.

Corollary 6.27. *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 6.28. *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 6.27 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 6.7. Use Proposition 6.26 to handle item 4.

Exercise 6.6. Prove Corollary 6.28.

Proposition 6.29 (Closure under sups, infs and limits). *Suppose that $(X,\mathcal{M})$ is a measurable space and $f_j : (X,\mathcal{M}) \to \overline{\mathbb{R}}$ for $j \in \mathbb{N}$ is a sequence of $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. Then*
$$\sup_j f_j, \quad \inf_j f_j, \quad \limsup_{j\to\infty} f_j \text{ and } \liminf_{j\to\infty} f_j$$
are all $\mathcal{M}/\mathcal{B}_{\overline{\mathbb{R}}}$ – measurable functions. (Note that this result is in generally false when $(X,\mathcal{M})$ is a topological space and measurable is replaced by continuous in the statement.)

Proof. Define $g_+(x) := \sup_j f_j(x)$, then
$$\begin{aligned}\{x : g_+(x) \le a\} &= \{x : f_j(x) \le a \ \forall j\} \\ &= \cap_j \{x : f_j(x) \le a\} \in \mathcal{M}\end{aligned}$$
so that g_+ is measurable. Similarly if $g_-(x) = \inf_j f_j(x)$ then
$$\{x : g_-(x) \ge a\} = \cap_j \{x : f_j(x) \ge a\} \in \mathcal{M}.$$
Since
$$\limsup_{j\to\infty} f_j = \inf_n \sup\{f_j : j \ge n\} \text{ and}$$
$$\liminf_{j\to\infty} f_j = \sup_n \inf\{f_j : j \ge n\}$$
we are done by what we have already proved. ■

Definition 6.30. *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 6.31. *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 6.29 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ■

Definition 6.32. *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
$$\varphi = \sum_{i=1}^n \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{6.5}$$
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
$$\varphi = \sum_{y \in \mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{6.6}$$

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

Theorem 6.33 (Approximation Theorem). *Let $f : X \to [0,\infty]$ be measurable and define, see Figure 6.3,*
$$\begin{aligned}\varphi_n(x) &:= \sum_{k=0}^{n2^n-1} \frac{k}{2^n} 1_{f^{-1}\left(\left(\frac{k}{2^n},\frac{k+1}{2^n}\right]\right)}(x) + n1_{f^{-1}((n2^n,\infty])}(x) \\ &= \sum_{k=0}^{n2^n-1} \frac{k}{2^n} 1_{\left\{\frac{k}{2^n} < f \le \frac{k+1}{2^n}\right\}}(x) + n1_{\{f > n2^n\}}(x)\end{aligned}$$
then $\varphi_n \le f$ for all n, $\varphi_n(x) \uparrow f(x)$ for all $x \in X$ and $\varphi_n \uparrow f$ uniformly on the sets $X_M := \{x \in X : f(x) \le M\}$ with $M < \infty$.

Moreover, if $f : X \to \mathbb{C}$ is a measurable function, then there exists simple functions φ_n such that $\lim_{n\to\infty} \varphi_n(x) = f(x)$ for all x and $|\varphi_n| \uparrow |f|$ as $n \to \infty$.

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

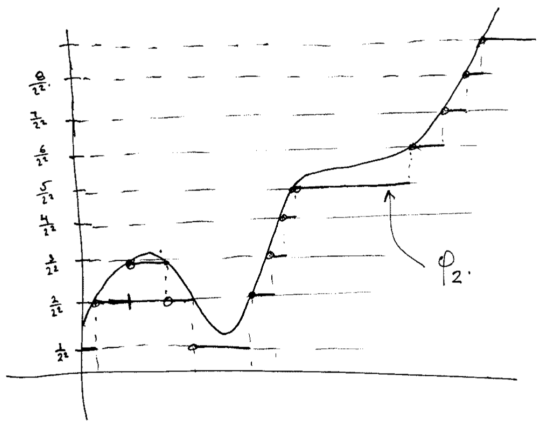


Fig. 6.3. Constructing simple functions approximating a function, $f : X \to [0, \infty]$.

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 $|u_n| \uparrow |\operatorname{Re} f|$, $|v_n| \uparrow |\operatorname{Im} f|$, $u_n \to \operatorname{Re} f$ and $v_n \to \operatorname{Im} f$ as $n \to \infty$. Let $\varphi_n = u_n + iv_n$, then

$$|\varphi_n|^2 = u_n^2 + v_n^2 \uparrow |\operatorname{Re} f|^2 + |\operatorname{Im} f|^2 = |f|^2$$

and $\varphi_n = u_n + iv_n \to \operatorname{Re} f + i \operatorname{Im} f = f$ as $n \to \infty$. ∎

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, the $\sum_{n=1}^\infty A_n \in \mathcal{L}$. (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 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. ∎

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 $A \cap B \notin \mathcal{L}$.

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

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

for the first case and

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

for the second case. ∎

Definition 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 $(X_1, \dots, X_n) \stackrel{d}{=} (Y_1, \dots, Y_n)$ if $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ have the **same distribution**, i.e. if*

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

More generally, if $\{X_i\}_{i=1}^\infty$ and $\{Y_i\}_{i=1}^\infty$ are two sequences of random variables on two probability spaces, $(\Omega, \mathcal{B}, P)$ and $(X, \mathcal{F}, Q)$ we write $\{X_i\}_{i=1}^\infty \stackrel{d}{=} \{Y_i\}_{i=1}^\infty$ iff $(X_1, \dots, X_n) \stackrel{d}{=} (Y_1, \dots, Y_n)$ for all $n \in \mathbb{N}$.

Exercise 7.1. 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, $S_n := X_1 + \dots + X_n$ and $T_n := Y_1 + \dots + Y_n$. Prove the following assertions.

1. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^k$ is a $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}^k}$ – measurable function, then $f(X_1, \dots, X_n) \stackrel{\mathrm{d}}{=} f(Y_1, \dots, Y_n)$.
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,
   $$\left\{\limsup_{n\to\infty} X_n \ge x\right\} = \{X_n \ge x \text{ i.o.}\},$$
   show
   $$P\left(\limsup_{n\to\infty} X_n \ge x\right) = \lim_{n\to\infty} \lim_{m\to\infty} P\left(\cup_{k=n}^m \{X_k \ge x\}\right).$$
   To use this identity you will also need to find $B \in \mathcal{B}_{\mathbb{R}^m}$ such that
   $$\cup_{k=n}^m \{X_k \ge x\} = \{(X_1, \dots, X_m) \in B\}.$$

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 σ – algebra. ∎

Exercise 7.2. 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
$$\begin{aligned} &B \setminus A \subset \cup_i (B_i \setminus A_i) \subset \cup_i A_i \triangle B_i \text{ and} \\ &A \setminus B \subset \cup_i (A_i \setminus B_i) \subset \cup_i A_i \triangle B_i \end{aligned}$$
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 (7.2). Since $\mathcal{L}$ contains the π – system, $\mathcal{A}$ it suffices by the $\pi - \lambda$ theorem to show $\mathcal{L}$ is a λ – system. Clearly, $\Omega \in \mathcal{L}$ since $\Omega \in \mathcal{A} \subset \mathcal{L}$. If $B_1 \subset B_2$ with $B_i \in \mathcal{L}$ and $\varepsilon > 0$, there exists $A_i \in \mathcal{A}$ such that $P(B_i \triangle A_i) = \mathbb{E}|1_{A_i} - 1_{B_i}| < \varepsilon/2$ 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

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

where $A := \cup_n A_n$. Hence it follows for N large enough that $P(B \triangle A^N) < \varepsilon$.

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 $J \subset \{1, 2, \dots, n\}$.

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

$$P\left(\cap_{j=1}^n A_j\right) = \prod_{j=1}^n P(A_j) \text{ for all } (A_1, \dots, A_n) \in \mathcal{C}_1 \times \dots \times \mathcal{C}_n.$$

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 $j = 2, 3, \dots, n$. We will begin by showing that

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

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

$$Q(A) = \frac{P(A \cap A_2 \cap \dots \cap A_n)}{P(A_2) \dots P(A_n)} = \frac{P(A \cap A_2 \cap \dots \cap A_n)}{P(A_2 \cap \dots \cap A_n)}$$
$$= P(A|A_2 \cap \dots \cap A_n) \text{ for all } A \in \sigma(\mathcal{C}_1).$$

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 $(A_2, \dots, A_n) \in \mathcal{C}_2 \times \dots \times \mathcal{C}_n$ were arbitrary we may now conclude that $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ are independent.

By applying the result we have just proved to the sequence, $\mathcal{C}_2, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ shows that $\sigma(\mathcal{C}_2), \mathcal{C}_3, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ are independent. Similarly we show inductively that

$$\sigma(\mathcal{C}_j), \mathcal{C}_{j+1}, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_{j-1})$$

are independent for each $j = 1, 2, \dots, n$. 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*

$$P(\cap_{t \in \Lambda} A_t) = \prod_{t \in \Lambda} P(A_t)$$

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 $\{\sigma(\mathcal{C}_t)\}_{t \in T}$ are independent as well.*

Example 7.16. Suppose that $\Omega = \Lambda^n$ where Λ is a finite set, $\mathcal{B} = 2^\Omega$, $P(\{\omega\}) = \prod_{j=1}^n q_j(\omega_j)$ where $q_j : \Lambda \to [0,1]$ are functions such that $\sum_{\lambda \in \Lambda} q_j(\lambda) = 1$. Let $\mathcal{C}_i := \left\{\Lambda^{i-1} \times A \times \Lambda^{n-i} : A \subset \Lambda\right\}$. Then $\{\mathcal{C}_i\}_{i=1}^n$ are independent. Indeed, if $B_i := \Lambda^{i-1} \times A_i \times \Lambda^{n-i}$, then

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

and we have

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

while

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

Definition 7.17. *A collections of random variables, $\{X_t : t \in T\}$ are* ***independent*** *iff $\{\sigma(X_t) : t \in T\}$ 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. $$P(\cap_{t \in \Lambda} \{X_t \in A_t\}) = \prod_{t \in \Lambda} P(X_t \in A_t)$$
   for all finite subsets, $\Lambda \subset T$, and all $A_t \in \mathcal{B}_\mathbb{R}$ for $t \in \Lambda$.
3. $$P(\cap_{t \in \Lambda} \{X_t \leq x_t\}) = \prod_{t \in \Lambda} P(X_t \leq x_t)$$
   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 $\sigma(X_t) = \{\{X_t \in A\} : A \in \mathcal{B}_{\mathbb{R}}\}$. Clearly 2. implies 3. holds. Finally, 3. implies 2. is an application of Corollary 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 $\sigma(\mathcal{C}_t) = \sigma(X_t)$. ■

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

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

for all $f_i : \Lambda \to \mathbb{R}$. Taking $A_i \subset \Lambda$ and $f_i := 1_{A_i}$ in the above identity shows that

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

as desired.

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

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

■

Definition 7.21. *As sequences of random variables, $\{X_n\}_{n=1}^\infty$, on a probability space, $(\Omega, \mathcal{B}, P)$, are* ***iid (= independent and identically distributed)*** *if they are independent and $(X_n)_* P = (X_k)_* P$ for all k, n. That is we should have*

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

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

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

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*

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

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

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

$$T_i\left((a,b]\right) := \left(a + \sigma_{i-1}\left(b-a\right), a + \sigma_i\left(b-a\right)\right].$$

Given $i_1, i_2, \dots, i_k \in \{0, 1, 2, \dots, n\}$, let

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

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 6.21 shows the functions, $X_k : (0,1] \to \mathbb{R}$ are measurable.

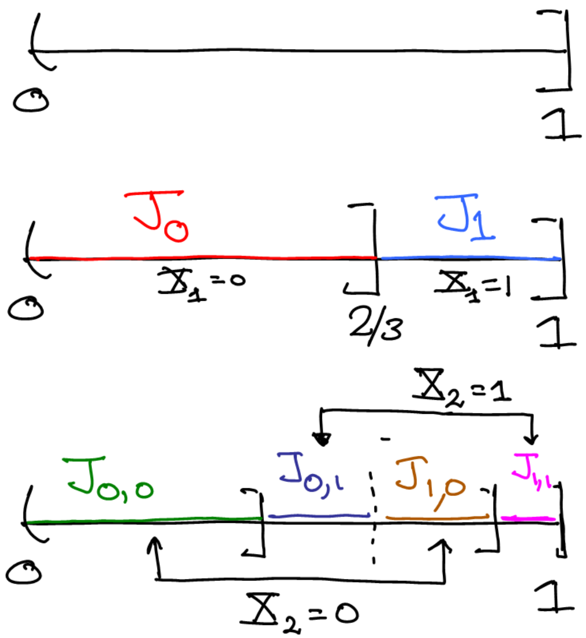


Fig. 7.1. Here we suppose that $p_0 = 2/3$ and $p_1 = 1/3$ and then we construct J_l and $J_{l,k}$ for $l, k \in \{0,1\}$.

Observe that

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

and so by induction,

$$m\left(J_{i_1,i_2,\dots,i_k}\right) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}.$$

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

$$m\left(\{X_1 = i_1, \dots, X_k = i_i\}\right) = m\left(J_{i_1,i_2,\dots,i_k}\right) = q_{i_k} q_{i_{k-1}} \dots q_{i_1}$$

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 $\mathcal{B} := \sigma\left(Y_1, Y_2, \dots, Y_n, \dots\right)$. Then there exists a unique probability measure, $P : \mathcal{B} \to [0,1]$ such that*

$$P\left(\{Y_1 = i_1, \dots, Y_k = i_i\}\right) = q_{i_1} \dots q_{i_k}.$$

Proof. Let $\{X_i\}_{i=1}^n$ be as in Theorem 7.22 and define $T : (0,1] \to \Omega$ by

$$T(x) = \left(X_1(x), X_2(x), \dots, X_k(x), \dots\right).$$

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

$$\begin{aligned} P\left(\{Y_1 = i_1, \dots, Y_k = i_i\}\right) &= m\left(T^{-1}\left(\{Y_1 = i_1, \dots, Y_k = i_i\}\right)\right) \\ &= m\left(\{Y_1 \circ T = i_1, \dots, Y_k \circ T = i_i\}\right) \\ &= m\left(\{X_1 = i_1, \dots, X_k = i_i\}\right) = q_{i_1} \dots q_{i_k}. \end{aligned}$$

■

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 5.26 or Theorem 7.22 fit the bill. ■

Proposition 7.25. *Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of iid 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

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

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

$$P\left(U < x + 2^{-(n+1)}\right) = P(U < x) + P\left(x \le U < x + 2^{-(n+1)}\right)$$
$$= x + P\left(x \le U < x + 2^{-(n+1)}\right).$$

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 σ – 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 π – systems. Therefore $\{\mathcal{B}_{T_s} = \sigma(\mathcal{C}_s)\}_{s \in S}$ is an independent family of σ – 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 iid 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 σ – algebras and hence $\{U_i\}_{i=1}^\infty$ is a sequence of iid. random variables with the uniform distribution.

Finally, let $F_i(x) := \mu((-\infty, x])$ for all $x \in \mathbb{R}$ and let $G_i(y) = \inf\{x : F_i(x) \ge y\}$. 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 iid with common continuous distribution function, F. In this case we have, for any $i \neq j$, that

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

This may be proved directly with some work or will be an easy consequence of Fubini's theorem to be considered later. 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 iid 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

$$1 = P(\Omega) = \sum_\sigma P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}).$$

Since $P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\})$ is independent of σ we may now conclude that

$$P(\{X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}\}) = \frac{1}{n!}$$

for all σ. As observed before the statement of the theorem, to each realization $(\varepsilon_1, \dots, \varepsilon_n)$, (here $\varepsilon_i \in \mathbb{N}$ with $\varepsilon_i \le i$) of $(R_1, \dots, R_n)$ there is a permutation, $\sigma = \sigma(\varepsilon_1, \dots, \varepsilon_n)$ such that $X_{\sigma 1} < X_{\sigma 2} < \dots < X_{\sigma n}$. From this it follows that

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

and therefore,

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

Since

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

we have shown that

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

■

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

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

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* $\varphi \in C^2((a,b) \to \mathbb{R})$ *with* $\varphi''(x) \ge 0$ *for all* $x \in (a,b)$. *Then* φ *satisfies;*

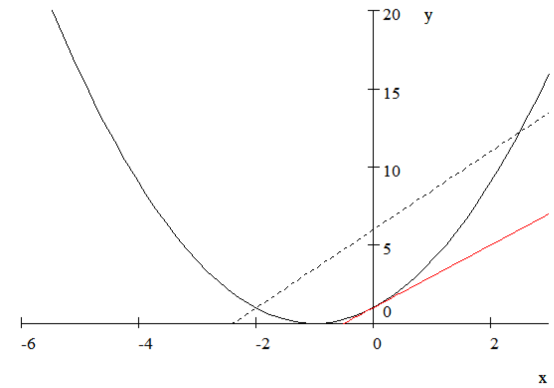


Fig. 7.2. A convex function, φ, along with a cord and a tangent line. Notice that the tangent line is always below φ and the cord lies above φ between the points of intersection of the cord with the graph of φ.

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

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

and

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

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

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 $f''(x) \ge 0$. Hence it follows by the mean value theorem that $f'(x) \ge 0$ for $x > x_0$ and $f'(x) \le 0$ for $x < x_0$ and therefore, $f(x) \ge 0$ 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 $\ddot{f}(t) = -(v-u)^2 \varphi''(u + t(v-u)) \le 0$. By the mean value theorem, there exists, $t_0 \in (0,1)$ such that $\dot{f}(t_0) = 0$ and then again by the mean value theorem, it follows that $\dot{f}(t) \le 0$ for $t > t_0$ and $\dot{f}(t) \ge 0$ for $t < t_0$. In particular $f(t) \ge f(1) = 0$ 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

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

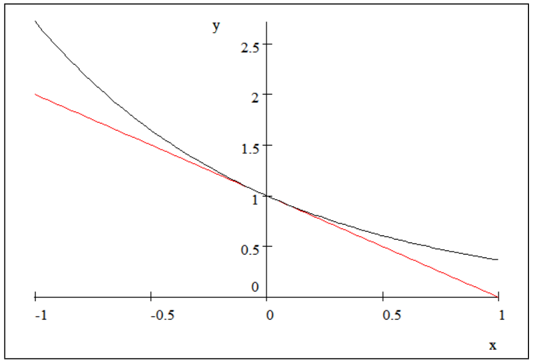


Fig. 7.3. A graph of $1 - x$ and e^{-x} showing that $1 - x \le e^{-x}$ for all x.

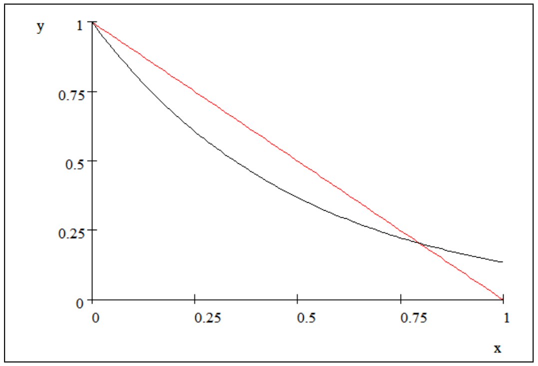


Fig. 7.4. A graph of $1 - x$ and e^{-2x} showing that $1 - x \ge e^{-2x}$ for all $x \in [0, 1/2]$.

Exercise 7.3. 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 (7.3). 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,

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

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 iid 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

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

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

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

or equivalently,

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

That is to say,

$$\limsup_{n\to\infty} \frac{X_n}{c_n} \leq \alpha \text{ a.s.}$$

and hence working as above,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} \leq 1\right) = P\left(\cap_{k=1}^{\infty} \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \leq \alpha_k\right\}\right) = 1.$$

Hence,

$$P\left(\limsup_{n\to\infty} \frac{X_n}{c_n} = 1\right) = P\left(\left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \geq 1\right\} \cap \left\{\limsup_{n\to\infty} \frac{X_n}{c_n} \leq 1\right\}\right) = 1.$$

■

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

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

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

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

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

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

which is convergent iff $\alpha > 1$. So by Proposition 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 iid distributed by the Poisson distribution with intensity, λ, i.e.

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

In this case we have

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

and

$$\begin{aligned} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=n}^{\infty} \frac{n!}{k!} \lambda^{k-n} \\ &= \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{n!}{(k+n)!} \lambda^k \le \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k = \frac{\lambda^n}{n!}. \end{aligned}$$

Thus we have shown that

$$\frac{\lambda^n}{n!} e^{-\lambda} \le P(X_1 \ge n) \le \frac{\lambda^n}{n!}.$$

Thus in terms of convergence issues, we may assume that

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

wherein we have used Stirling's formula,

$$x! \sim \sqrt{2\pi x} e^{-x} x^x.$$

Now suppose that we wish to choose c_n so that

$$P(X_1 \ge c_n) \sim 1/n.$$

This suggests that we need to solve the equation, $x^x = n$. Taking logarithms of this equation implies that

$$x = \frac{\ln n}{\ln x}$$

and upon iteration we find,

$$\begin{aligned} x &= \frac{\ln n}{\ln\left(\frac{\ln n}{\ln x}\right)} = \frac{\ln n}{\ell_2(n) - \ell_2(x)} = \frac{\ln n}{\ell_2(n) - \ell_2\left(\frac{\ln n}{\ln x}\right)} \\ &= \frac{\ln n}{\ell_2(n) - \ell_3(n) + \ell_3(x)}. \end{aligned}$$

where $\ell_k = \overbrace{\ln \circ \ln \circ \dots \circ \ln}^{k\text{ - times}}$. Since, $x \le \ln(n)$, it follows that $\ell_3(x) \le \ell_3(n)$ 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

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

Since

$$\ln(\lambda/e)^{\alpha c_n} = \alpha c_n \ln(\lambda/e) = \alpha \frac{\ln n}{\ell_2(n)} \ln(\lambda/e) = \ln n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}},$$

it follows that

$$(\lambda/e)^{\alpha c_n} = n^{\alpha \frac{\ln(\lambda/e)}{\ell_2(n)}}.$$

Therefore,

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

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, $(\Omega, \mathcal{B})$ with $\mathcal{B} = \sigma(X_1, X_2, \dots)$. Let $\mathcal{B}_n := \sigma(X_1, \dots, X_n)$, $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \dots)$, and we will call, $\mathcal{T} := \cap_{n=1}^{\infty} \mathcal{T}_n$, the **tail** σ – **field** and events, $A \in \mathcal{T}$, are called **tail events.**

Example 7.38. Here are some example of tail events and tail measurable random variables:

1. $\{\sum_{n=1}^{\infty} X_n \text{ converges}\} \in \mathcal{T}$.
2. both $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are $\mathcal{T}$ – measurable.
3. $\{\lim X_n \text{ exists}\} = \{\limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n\} \in \mathcal{T}$.
4. Let $S_n := X_1 + \cdots + X_n$, then $\left\{\lim_{n\to\infty} \frac{S_n}{n} = 0\right\} \in \mathcal{T}$. Indeed, for any $k \in \mathbb{N}$,
$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{(X_{k+1} + \cdots + X_n)}{n}$$
from which it follows that $\left\{\lim_{n\to\infty} \frac{S_n}{n} = 0\right\} \in \mathcal{T}_k$ 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 in under the added condition that $P(X \in \mathbb{R}) = 1$.

For each $x \in \mathbb{R}$, $\{X \leq x\} \in \mathcal{F}$ and therefore, $P(X \leq x)$ is either 0 or 1. Since the function, $f(x) := P(X \leq x) \in \{0, 1\}$ 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 $A \in \mathcal{T}$. In particular the tail events in Example 7.38 have probability either* 0 *or* 1 *and* $\limsup_{n\to\infty} X_n$ *and* $\liminf_{n\to\infty} X_n$ *are both constant a.s.*

Proof. Let $A \in \mathcal{T}$. 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 $\sigma(\cup \mathcal{B}_n) = \vee_{n=1}^{\infty} \mathcal{B}_n$. 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. ■

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 $T_\pi(\omega) = (\omega_{\pi 1}, \omega_{\pi 2}, \dots)$.

Definition 7.42. *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 7.43 (Hewitt-Savage Zero-One Law). *Let P be a probability measure on $(\Omega, \mathcal{B})$ such that $\{X_n\}_{n=1}^{\infty}$ is an iid 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, 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

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

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 7.44 (Some Random Walk $0-1$ Law Results). Continue the notation in Proposition 7.43.

1. As above, if $S_n = X_1 + \dots + X_n$, then $P(S_n \in B \text{ i.o.}) \in \{0, 1\}$ for all $B \in \mathcal{B}_{\mathbb{R}}$. Indeed, if π is a finite permutation,

$$T_\pi^{-1}(\{S_n \in B \text{ i.o.}\}) = \{S_n \circ T_\pi \in B \text{ i.o.}\} = \{S_n \in B \text{ i.o.}\}.$$

Hence $\{S_n \in B \text{ i.o.}\}$ is in the permutation invariant σ – field. The same goes for $\{S_n \in B \text{ a.a.}\}$

2. If $P(X_1 \neq 0) > 0$, then $\limsup_{n\to\infty} S_n = \infty$ a.s. or $\limsup_{n\to\infty} S_n = -\infty$ a.s. Indeed,

$$T_\pi^{-1}\left\{\limsup_{n\to\infty} S_n \le x\right\} = \left\{\limsup_{n\to\infty} S_n \circ T_\pi \le x\right\} = \left\{\limsup_{n\to\infty} S_n \le x\right\}$$

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 $P(X_1 \neq 0) > 0$ and $X_1 \overset{d}{=} -X_1$, 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 iid and $-X_n \overset{d}{=} X_n$, it follows that $\{X_n\}_{n=1}^\infty \overset{d}{=} \{-X_n\}_{n=1}^\infty$.The results of Exercise 7.1 then imply that $\limsup_{n\to\infty} S_n \overset{d}{=} \limsup_{n\to\infty} (-S_n)$ 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

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

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 6.33 $\varphi_n \uparrow f$ 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,

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

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 8.33). 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

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

is a Riemann sum. Therefore if $\{\pi_k\}_{k=1}^\infty$ is a sequence of partitions with $\lim_{k\to\infty} \text{mesh}(\pi_k) = 0$, 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 $|f_{\pi_k}(x)| \le g(x) := M 1_{(a,b]}(x)$ for all $x \in (a,b]$ where $M := \max_{x\in[a,b]} |f(x)| < \infty$. Since $\int_{\mathbb{R}} g dm = M(b-a) < \infty$, we may apply D.C.T. to conclude,

$$\lim_{k\to\infty} \int_a^b f_{\pi_k} \ dm = \int_a^b \lim_{k\to\infty} f_{\pi_k} \ dm = \int_a^b f \ dm.$$

This equation with Eq. (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 8.48 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 $\left|1_{(a,x)}(y) f(y)\right| \le 1_{(a,b)}(y) |f(y)|$ is an L^1 – function, it follows from the dominated convergence Theorem 8.33 that F is continuous on $[a,b]$. Simple manipulations show,

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

and the latter expression, by the continuity of f, goes to zero as $h \to 0$. This shows $F' = f$ on (a,b).

For the converse direction, we have by assumption that $G'(x) = F'(x)$ for $x \in (a,b)$. Therefore by the mean value theorem, $F - G = C$ for some constant C. Hence

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

■

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

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

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

Therefore, for $x < n$, we have

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

from which it follows that

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

From Example 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 $0 \le y < 1$,

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

Since $h(0) = 0$ and

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

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

Example 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 $L^+ = L^+(\mathcal{B}) = \{f : X \to [0,\infty] : f$ 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 $\lambda \ge 0$, 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 $0 \le f \le g$, then*
$$\int_X f d\mu \le \int_X g d\mu. \tag{8.2}$$
3. *For all $\varepsilon > 0$ and $p > 0$,*
$$\mu(f \ge \varepsilon) \le \frac{1}{\varepsilon^p}\int_X f^p 1_{\{f\ge\varepsilon\}} d\mu \le \frac{1}{\varepsilon^p}\int_X f^p d\mu. \tag{8.3}$$
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,

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

2. Since

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

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

3. Since $1_{\{f\ge\varepsilon\}} \le 1_{\{f\ge\varepsilon\}}\frac{1}{\varepsilon}f \le \frac{1}{\varepsilon}f$ we have

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

and by monotonicity and the multiplicative property of the integral,

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

4. If $\mu(f = \infty) > 0$, then $\varphi_n := n 1_{\{f=\infty\}}$ is a simple function such that $\varphi_n \le f$ 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 $\varphi = \sum_{z\in[0,\infty)} z 1_{\{\varphi=z\}}$ and

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

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

$$\sum_\Lambda f\rho \le \int_X f d\mu.$$

Since Λ is arbitrary, this implies

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

■

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 $f_n \le f_m \le f$, for all $n \le m < \infty$,

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

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

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

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

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

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

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

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

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.

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

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

■

Remark 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 6.33 which relies heavily on the measurability of the functions to be approximated.

Proposition 8.18. *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 8.19. *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 8.18,

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

■

Lemma 8.20 (Fatou's Lemma). *If $f_n : X \to [0,\infty]$ is a sequence of measurable functions then*

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

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

$$\int g_k \leq \int f_n \text{ for all } n \geq k$$

and therefore

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

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

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

■

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

Lemma 8.21 (The First Borell – Carntelli Lemma). *Let $(X, \mathcal{B}, \mu)$ be a measure space, $A_n \in \mathcal{B}$, and set*

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

If $\sum_{n=1}^{\infty} \mu(A_n) < \infty$ then $\mu(\{A_n \text{ i.o.}\}) = 0$.

Proof. (First Proof.) Let us first observe that

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

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 8.22. *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 8.23. 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,

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

we find

$$\int_{[0,1]} f(x) dm(x) = \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} \frac{1}{\sqrt{|x - r_n|}} dx \leq \sum_{n=1}^{\infty} 2^{-n} 4 = 4 < \infty.$$

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 8.24. *A measurable function $f : X \to \bar{\mathbb{R}}$ is **integrable** if $f_+ := f 1_{\{f\geq 0\}}$ and $f_- = -f\ 1_{\{f\leq 0\}}$ are **integrable**. We write $L^1(\mu; \mathbb{R})$ for the space of real valued integrable functions. For $f \in L^1(\mu; \mathbb{R})$, 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 8.25 (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 8.26. 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 8.18 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 8.27. *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 8.18. ■

Definition 8.28. *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 8.29. *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 8.18,

$$\left|\int_X f d\mu\right| = \int_X \mathrm{Re}\left(e^{-i\theta} f\right) d\mu \le \int_X \left|\mathrm{Re}\left(e^{-i\theta} f\right)\right| d\mu \le \int_X |f| \, d\mu.$$

■

Proposition 8.30. *Let $f, g \in \mathrm{L}^1(\mu)$, then*

1. *The set $\{f \ne 0\}$ is σ – finite, in fact $\{|f| \ge \frac{1}{n}\} \uparrow \{f \ne 0\}$ and $\mu(|f| \ge \frac{1}{n}) < \infty$ for all n.*
2. *The following are equivalent*
   a) $\int_E f = \int_E g$ *for all* $E \in \mathcal{B}$
   b) $\int\limits_X |f - g| = 0$
   c) $f = g$ *a.e.*

Proof. 1. By Chebyshev's inequality, Lemma 8.13,

$$\mu(|f| \ge \frac{1}{n}) \le n \int_X |f| \, d\mu < \infty$$

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 $1_E \mathrm{Re}(f - g) \ge 0$, we learn that

$$0 = \mathrm{Re}\int_E (f - g) d\mu = \int 1_E \mathrm{Re}(f - g) \Longrightarrow 1_E \mathrm{Re}(f - g) = 0 \text{ a.e.}$$

This implies that $1_E = 0$ a.e. which happens iff

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

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

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

■

Definition 8.31. *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 8.32. 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 8.33 (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 $f \in L^1(\mu)$. 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 $\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$, we have shown,

$$\int_X g d\mu \pm \int_X f d\mu \le \int_X g d\mu + \begin{cases}\liminf_{n\to\infty}\int_X f_n d\mu \\ -\limsup_{n\to\infty}\int_X f_n d\mu\end{cases}$$

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 $\lim_{n\to\infty}\int_X f_n d\mu$ exists and is equal to $\int_X f d\mu$. ■

Exercise 8.4. Give another proof of Proposition 8.29 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 6.33 along with the dominated convergence Theorem 8.33 to handle the general case.

Corollary 8.34. *Let $\{f_n\}_{n=1}^\infty \subset L^1(\mu)$ be a sequence such that $\sum_{n=1}^\infty \|f_n\|_{L^1(\mu)} < \infty$, 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 $\sum_{n=1}^\infty \|f_n\|_{L^1(\mu)} < \infty$ is equivalent to $\sum_{n=1}^\infty |f_n| \in L^1(\mu)$. Hence $\sum_{n=1}^\infty f_n$ is almost everywhere convergent and if $S_N := \sum_{n=1}^N f_n$, then

$$|S_N| \le \sum_{n=1}^N |f_n| \le \sum_{n=1}^\infty |f_n| \in L^1(\mu).$$

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 8.35 (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 8.34 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 8.36 (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. (The proof is essentially the same as for sums.) 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 $f(t,\cdot) \in \mathrm{L}^1(\mu)$ 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 8.37. 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 $\lambda \ge 2\varepsilon > 0$ and $n \in \mathbb{N}$ there exists $C_n(\varepsilon) < \infty$ such that

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

Using this fact, Corollary 8.36 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 8.38. Corollary 8.36 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 8.36 to $g(t) := \int_0^\infty 1_{x\le t} dm(x)$,

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

The last integral is zero since $\frac{\partial}{\partial t} 1_{x\le t} = 0$ unless $t = x$ in which case it is not defined. On the other hand $g(t) = t$ so that $\dot{g}(t) = 1$. (The reader should decide which hypothesis of Corollary 8.36 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 $0 \leq \varphi_n \uparrow f$. 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,

$$\begin{aligned}\int_X f d\nu &= \int_X f_+ d\nu - \int_X f_- d\nu = \int_X f_+ \rho d\mu - \int_X f_- \rho d\mu \\ &= \int_X [f_+ \rho - f_- \rho] \, d\mu = \int_X f \rho d\mu.\end{aligned}$$

The complex case easily follows from this identity.

Notation 8.39 *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.

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,

$$\int_\mathbb{R} h \circ F \cdot F' dm = \int_\mathbb{R} h dm \tag{8.11}$$

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. (8.11) 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

$$\begin{aligned}\int_\mathbb{R} h \circ F \cdot F' dm &= \int_\mathbb{R} h \circ F d\nu = \int_\mathbb{R} h \circ F d\left(F^{-1}_* m\right) = \int_\mathbb{R} (h \circ F) \circ F^{-1} dm \\ &= \int_\mathbb{R} h dm.\end{aligned}$$

This result is also valid for all $h \in L^1(m)$.

Lemma 8.40. *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

$$P(X \geq x) \leq \frac{1}{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{8.12}$$

and[1]

$$\lim_{x\to\infty} \frac{P(X \geq x)}{\frac{1}{x}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}} = 1. \tag{8.13}$$

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. (8.12) follows. To prove Eq. (8.13), let $\alpha > 1$, then

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

Hence

$$\frac{P(X \geq x)}{\frac{1}{x}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}} \geq \frac{\int_x^{\alpha x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy}{\frac{1}{x}\frac{1}{\sqrt{2\pi}} e^{-x^2/2}} \geq \frac{1}{\alpha} \left[\frac{e^{-x^2/2} - e^{-\alpha^2 x^2/2}}{e^{-x^2/2}}\right] = \frac{1}{\alpha}\left[1 - e^{-(\alpha^2-1)x^2/2}\right].$$

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. (8.12) 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 8.41. Let $\{X_n\}_{n=1}^\infty$ be iid standard normal random variables. Then

$$P(X_n \geq \alpha c_n) \sim \frac{1}{\alpha c_n} e^{-\alpha^2 c_n^2/2}.$$

Now, suppose that we take c_n so that

$$e^{-c_n^2/2} = \frac{C}{n}$$

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

$$\limsup_{n\to\infty} \frac{X_n}{\sqrt{2\ln(n)}} = 1 \text{ a.s.}.$$

[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

8.5 Measurability on Complete Measure Spaces

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

Proposition 8.42. *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,

$$(1_E g)^{-1}(A) = \begin{cases} E^c \cup (1_E g)^{-1}(A \setminus \{0\}) & \text{if } 0 \in A \\ (1_E g)^{-1}(A) & \text{if } 0 \notin A \end{cases} \tag{8.14}$$

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. (8.14) shows that $1_E g$ is measurable. 2. Let $E = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}$ by assumption $E \in \mathcal{B}$ and $\mu(E) = 0$. Since $g := 1_E f = \lim_{n\to\infty} 1_{E^c} f_n$, g is measurable. Because $f = g$ on E^c and $\mu(E) = 0$, $f = g$ a.e. so by part 1. f is also measurable. ■

The above results are in general false if $(X, \mathcal{B}, \mu)$ is not complete. For example, let $X = \{0, 1, 2\}$, $\mathcal{B} = \{\{0\}, \{1, 2\}, X, \varphi\}$ and $\mu = \delta_0$. Take $g(0) = 0$, $g(1) = 1$, $g(2) = 2$, then $g = 0$ a.e. yet g is not measurable.

Lemma 8.43. *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 8.42 that f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable. Conversely if f is $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable, by considering $f_\pm$ we may assume that $f \geq 0$. Choose $(\bar{\mathcal{M}}, \mathcal{B})$ – measurable simple function $\varphi_n \geq 0$ such that $\varphi_n \uparrow f$ as $n \to \infty$. Writing

$$\varphi_n = \sum a_k 1_{A_k}$$

with $A_k \in \bar{\mathcal{M}}$, we may choose $B_k \in \mathcal{M}$ such that $B_k \subset A_k$ and $\bar{\mu}(A_k \setminus B_k) = 0$. Letting

$$\tilde{\varphi}_n := \sum a_k 1_{B_k}$$

we have produced a $(\mathcal{M}, \mathcal{B})$ – measurable simple function $\tilde{\varphi}_n \geq 0$ such that $E_n := \{\varphi_n \neq \tilde{\varphi}_n\}$ has zero $\bar{\mu}$ – measure. Since $\bar{\mu}(\cup_n E_n) \leq \sum_n \bar{\mu}(E_n)$, there exists $F \in \mathcal{M}$ such that $\cup_n E_n \subset F$ and $\mu(F) = 0$. It now follows that

$$1_F \cdot \tilde{\varphi}_n = 1_F \cdot \varphi_n \uparrow g := 1_F f \text{ as } n \to \infty.$$

This shows that $g = 1_F f$ is $(\mathcal{M}, \mathcal{B})$ – measurable and that $\{f \neq g\} \subset F$ has $\bar{\mu}$ – measure zero. Since $f = g$, $\bar{\mu}$ – a.e., $\int_X f d\bar{\mu} = \int_X g d\bar{\mu}$ so to prove Eq. (8.15) it suffices to prove

$$\int_X g d\bar{\mu} = \int_X g d\mu. \tag{8.15}$$

Because $\bar{\mu} = \mu$ on $\mathcal{M}$, Eq. (8.15) is easily verified for non-negative $\mathcal{M}$ – measurable simple functions. Then by the monotone convergence theorem and the approximation Theorem 6.33 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 $(\mathrm{Re}\, g)_\pm$ and $(\mathrm{Im}\, g)_\pm$. ■

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

$$\pi = \{a = t_0 < t_1 < \cdots < t_n = b\} \tag{8.16}$$

of $[a, b]$ let

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

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

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

[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.

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

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 8.44. *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 8.45. *If π' and π are two partitions of $[a,b]$ and $\pi \subset \pi'$ then*

$$G_\pi \geq G_{\pi'} \geq f \geq g_{\pi'} \geq g_\pi \text{ and}$$
$$S_\pi f \geq S_{\pi'} f \geq s_{\pi'} f \geq s_\pi f.$$

There exists an increasing sequence of partitions $\{\pi_k\}_{k=1}^\infty$ such that $\operatorname{mesh}(\pi_k) \downarrow 0$ 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

$$G := \lim_{k\to\infty} G_{\pi_k} \text{ and } g := \lim_{k\to\infty} g_{\pi_k} \tag{8.17}$$

then by the dominated convergence theorem,

$$\int_{[a,b]} g dm = \lim_{k\to\infty} \int_{[a,b]} g_{\pi_k} = \lim_{k\to\infty} s_{\pi_k} f = \underline{\int_a^b} f(x)dx \tag{8.18}$$

and

$$\int_{[a,b]} G dm = \lim_{k\to\infty} \int_{[a,b]} G_{\pi_k} = \lim_{k\to\infty} S_{\pi_k} f = \overline{\int_a^b} f(x)dx. \tag{8.19}$$

Notation 8.46 *For $x \in [a,b]$, let*

$$H(x) = \limsup_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \sup\{f(y) : |y-x| \leq \varepsilon,\ y \in [a,b]\} \text{ and}$$
$$h(x) = \liminf_{y\to x} f(y) := \lim_{\varepsilon\downarrow 0} \inf\{f(y) : |y-x| \leq \varepsilon,\ y \in [a,b]\}.$$

Lemma 8.47. *The functions $H, h : [a,b] \to \mathbb{R}$ satisfy:*

1. *$h(x) \leq f(x) \leq H(x)$ 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 $\operatorname{mesh}(\pi_k) \downarrow 0$ and G and g are defined as in Eq. (8.17), then*
$$G(x) = H(x) \geq f(x) \geq h(x) = g(x) \quad \forall\ x \notin \pi := \cup_{k=1}^\infty \pi_k. \tag{8.20}$$
(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 $h(x) \leq f(x) \leq H(x)$ 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 $x \notin \pi$,
$$G_k(x) \geq H(x) \geq f(x) \geq h(x) \geq g_k(x)\ \forall\ k$$
and letting $k \to \infty$ in this equation implies
$$G(x) \geq H(x) \geq f(x) \geq h(x) \geq g(x)\ \forall\ x \notin \pi. \tag{8.21}$$
Moreover, given $\varepsilon > 0$ and $x \notin \pi$,
$$\sup\{f(y) : |y-x| \leq \varepsilon,\ y \in [a,b]\} \geq G_k(x)$$
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 $\sup_{|y-x|\leq\varepsilon} f(y) \geq G(x)$ and therefore, that
$$H(x) = \limsup_{y\to x} f(y) \geq G(x)$$
for all $x \notin \pi$. Combining this equation with Eq. (8.21) 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. (8.20) 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 8.48. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then*

$$\overline{\int_a^b} f = \int_{[a,b]} H dm \text{ and } \underline{\int_a^b} f = \int_{[a,b]} h dm \tag{8.22}$$

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

$$\int_{[a,b]} H dm = \int_a^b f(x)dx = \int_{[a,b]} f d\bar{m} = \int_{[a,b]} h dm. \tag{8.23}$$

Proof. Let $\{\pi_k\}_{k=1}^\infty$ be an increasing sequence of partitions of $[a,b]$ as described in Lemma 8.45 and let G and g be defined as in Lemma 8.47. Since $m(\pi) = 0$, $H = G$ a.e., Eq. (8.22) is a consequence of Eqs. (8.18) and (8.19). From Eq. (8.22), f is Riemann integrable iff

$$\int_{[a,b]} H dm = \int_{[a,b]} h dm$$

and because $h \le f \le H$ this happens iff $h(x) = H(x)$ for m - a.e. x. Since $E = \{x : H(x) \neq h(x)\}$, this last condition is equivalent to E being a m – null set. In light of these results and Eq. (8.20), the remaining assertions including Eq. (8.23) are now consequences of Lemma 8.43. ■

Notation 8.49 *In view of this theorem we will often write $\int_a^b f(x)dx$ for $\int_a^b f dm$.*

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

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta.$$

[3] f need not be Borel measurable.

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 8.45.

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

$$\begin{aligned}\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^4\right] &= \frac{1}{n^4}\left(n\gamma + 3n(n-1)\sigma^4\right) \\ &= \frac{1}{n^2}\left[n^{-1}\gamma + 3\left(1 - n^{-1}\right)\sigma^4\right]\end{aligned}$$

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 8.21 that $\lim_{n\to\infty} \frac{S_n}{n} = \mu$ a.s.