Bruce K. Driver

Math 280 (Probability Theory) Lecture Notes

October 7, 2009 *File:prob.tex*

Contents

Part

Homework Problems

-3

Math 280A Homework Problems Fall 2009

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

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

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

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

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

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

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

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

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

Math 280C Homework Problems Spring 2010

0

Math 286 Homework Problems Spring 2008

Part I

Background Material

1

Limsups, Liminfs and Extended Limits

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

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

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

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

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

while

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{``}=\text{''} \infty - \infty.$$

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

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

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

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

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

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

Therefore,

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

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

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

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

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

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

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

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

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

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

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

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

1. $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$ *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. 1. Since

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

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

Now suppose that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a \in \mathbb{R}$. 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.

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

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

while

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

Thus in this case we have

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

■

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

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

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

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

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

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

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

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

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

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

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

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

On the other hand,

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

so that

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

Second Proof. Let

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

and

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

Since

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

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

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

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

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

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

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

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

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

Then

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

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

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

Then

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

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

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

and

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

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

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

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

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

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

Then

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

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

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

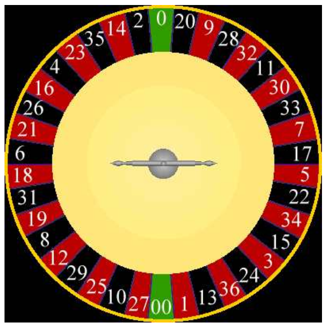
Exercise 1.2. Prove Proposition 1.10. **Hint:** apply the MCT by applying the monotone convergence theorem with $f_n(i) := \inf_{m\geq n} h_m(i)$.

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

2

Basic Probabilistic Notions

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



Example 2.2. 1. The sample space for flipping a coin one time could be taken to be, $\Omega = \{0, 1\}$.
2. The sample space for flipping a coin N -times could be taken to be, $\Omega = \{0, 1\}^N$ and for flipping an infinite number of times,
$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\}\} = \{0, 1\}^{\mathbb{N}}.$$
3. If we have a roulette wheel with 38 entries, then we might take
$$\Omega = \{00, 0, 1, 2, \dots, 36\}$$
for one spin,
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^N$$
for N spins, and
$$\Omega = \{00, 0, 1, 2, \dots, 36\}^{\mathbb{N}}$$
for an infinite number of spins.
4. If we throw darts at a board of radius R, we may take
$$\Omega = D_R := \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le R\right\}$$
for one throw,
$$\Omega = D_R^N$$
for N throws, and
$$\Omega = D_R^{\mathbb{N}}$$
for an infinite number of throws.
5. Suppose we release a perfume particle at location $x \in \mathbb{R}^3$ and follow its motion for all time, $0 \le t < \infty$. In this case, we might take,
$$\Omega = \left\{\omega \in C\left([0, \infty), \mathbb{R}^3\right) : \omega(0) = x\right\}.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

Since

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

we see that

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

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

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

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

The unfortunate fact is that we can not always assign a desired probability function, $P\left(A\right)$, 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 $P\left(S\right) = 1$, P is invariant under rotations, and P is continuous under taking decreasing limits.*

Proof. We are going to use the fact proved below in Proposition 5.1, that the continuity condition on P is equivalent to the σ – additivity of P. For $z \in S$ and $N \subset S$ let

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

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

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

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

Let

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

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

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

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

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

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

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

Part II

Formal Development

3

Preliminaries

3.1 Set Operations

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

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

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

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

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

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

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

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

For $A \in 2^X$ let

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

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

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

We also define the symmetric difference of A and B by

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

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

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

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

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

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

and

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

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

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

and this may be expressed as

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

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

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

which may be written as

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

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

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

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

Lemma 3.6. *We have:*

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

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

Lemma 3.8 (Basic Properties of Countable Sets).

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

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

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

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

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

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

3.2 Exercises

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

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

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

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

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

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

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

3.3 Algebraic sub-structures of sets

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

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

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

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

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

Example 3.13. Here are some examples of algebras.

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

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

Proof. Simply take

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

and

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

■

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

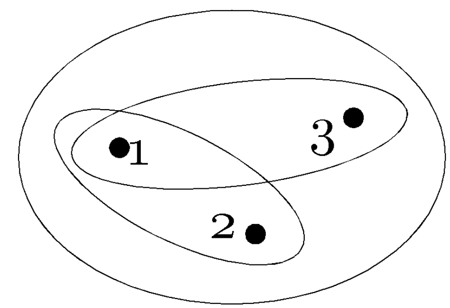


Fig. 3.1. A collection of subsets.

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

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

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

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

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

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

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

Hint: make use of Exercise 3.6.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

is also a semi-field.

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

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

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

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

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

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

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

4

Finitely Additive Measures / Integration

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

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

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

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

Proof.

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

■

4.1 Examples of Measures

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

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

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

Then

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

defines a measure on 2^{Ω}.

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

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

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

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

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

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

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

Example 4.6. Suppose that X is a 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 $A \subset X$.

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

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

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

$$\mu(A) = \sum_i \mu(E_i). \tag{4.5}$$

To prove existence, the main point is to show that $\mu(A)$ in Eq. (4.5) is well defined; i.e. if we also have $A = \sum_j F_j$ with $F_j \in \mathcal{S}$, then we must show

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

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. (4.6) holds. It is now easy to verify that μ extended to $\mathcal{A}$ as in Eq. (4.5) is an additive measure on $\mathcal{A}$. ■

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

$$\mathcal{S} = \{(a,b] \cap \mathbb{R} : -\infty \le a \le b \le \infty\}, \tag{4.7}$$

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

$$\mu((a,b] \cap \mathbb{R}) = F(b) - F(a) \;\forall\; a \le b \text{ in } \bar{\mathbb{R}}. \tag{4.8}$$

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. (4.8) holds. (Eventually we will only be interested in the case where $F(-\infty) = \lim_{a\downarrow -\infty} F(a)$ and $F(\infty) = \lim_{b\uparrow\infty} F(b)$.)

Proof. Given a finitely additive probability measure μ, let

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

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

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

Conversely, suppose $F : \bar{\mathbb{R}} \to [0,1]$ as in the statement of the theorem is given. Define μ on $\mathcal{S}$ using the formula in Eq. (4.8). The argument will be completed by showing μ is additive on $\mathcal{S}$ and hence, by Proposition 4.7, has a unique extension to a finitely additive measure on $\mathcal{A}$. Suppose that

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

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

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

Therefore, by the telescoping series argument,

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

■

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

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

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

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

Working as above we find,

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

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

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

4.2 Simple Random Variables

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

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

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

1. *$\mathbb{S}(\mathcal{A})$ is a sub-algebra of all functions from Ω to $\mathbb{C}$.*
2. *$f : \Omega \to \mathbb{C}$, is a $\mathcal{A}$ – simple random variable iff there exists $\alpha_i \in \mathbb{C}$ and $A_i \in \mathcal{A}$ for $1 \le i \le n$ for some $n \in \mathbb{N}$ such that*

$$f = \sum_{i=1}^{n} \alpha_i 1_{A_i}. \tag{4.9}$$

3. *For any function,* $F:\mathbb{C}\to\mathbb{C}$, $F\circ f\in\mathbb{S}(\mathcal{A})$ *for all* $f\in\mathbb{S}(\mathcal{A})$. *In particular,* $|f|\in\mathbb{S}(\mathcal{A})$ *if* $f\in\mathbb{S}(\mathcal{A})$.

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

$$\{f+cg=\lambda\}=\bigcup_{a,b\in\mathbb{C}:a+cb=\lambda}(\{f=a\}\cap\{g=b\})\in\mathcal{A} \tag{4.10}$$

and

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

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

2. Since $\mathbb{S}(\mathcal{A})$ is an algebra, every f of the form in Eq. (4.9) is in $\mathbb{S}(\mathcal{A})$. Conversely if $f\in\mathbb{S}(\mathcal{A})$ it follows by definition that $f=\sum_{\alpha\in f(\Omega)}\alpha 1_{\{f=\alpha\}}$ which is of the form in Eq. (4.9).

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

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

■

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

$$f=\sum_{i=1}^k\alpha_i 1_{B_i} \tag{4.12}$$

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

Corollary 4.13. *Suppose that Λ is a finite set and $Z:X\to\Lambda$ is a function. Let*

$$\mathcal{A}:=\mathcal{A}(Z):=Z^{-1}\left(2^\Lambda\right):=\left\{Z^{-1}(E):E\subset\Lambda\right\}.$$

Then $\mathcal{A}$ is an algebra and $f:X\to\mathbb{C}$ is an $\mathcal{A}$ – simple function iff $f=F\circ Z$ for some function $F:\Lambda\to\mathbb{C}$.

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

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

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

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

4.2.1 The algebraic structure of simple functions*

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

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

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

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

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

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

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

Exercise 4.2. Continuing the notation introduced above:

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

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

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

Solution to Exercise (4.2).

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

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

and

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

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

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

3. If $f:\Omega\to\mathbb{C}$ is a simple function such that $1_{\{f=\lambda\}}\in\mathbb{S}$ for all $\lambda\in\mathbb{C}$, then $f=\sum_{\lambda\in\mathbb{C}}\lambda 1_{\{f=\lambda\}}\in\mathbb{S}$. Conversely, by Lemma 4.15, if $f\in\mathbb{S}$ then $1_{\{f=\lambda\}}\in\mathbb{S}$ for all $\lambda\in\mathbb{C}$. Therefore, a simple function, $f:X\to\mathbb{C}$ is in $\mathbb{S}$ iff $1_{\{f=\lambda\}}\in\mathbb{S}$ for all $\lambda\in\mathbb{C}$. With this preparation, we are now ready to complete the verification.
   First off,
$$A\in\mathcal{A}(\mathbb{S}(\mathcal{A}))\iff 1_A\in\mathbb{S}(\mathcal{A})\iff A\in\mathcal{A}$$
   which shows that $\mathcal{A}(\mathbb{S}(\mathcal{A}))=\mathcal{A}$. Similarly,
$$\begin{aligned} f\in\mathbb{S}(\mathcal{A}(\mathbb{S}))&\iff\{f=\lambda\}\in\mathcal{A}(\mathbb{S})\ \forall\,\lambda\in\mathbb{C}\\ &\iff 1_{\{f=\lambda\}}\in\mathbb{S}\ \forall\,\lambda\in\mathbb{C}\\ &\iff f\in\mathbb{S}\end{aligned}$$
   which shows $\mathbb{S}(\mathcal{A}(\mathbb{S}))=\mathbb{S}$.

4.3 Simple Integration

Definition 4.16 (Simple Integral). *Suppose now that P is a finitely additive probability measure on an algebra $\mathcal{A}\subset 2^X$. For $f\in\mathbb{S}(\mathcal{A})$ the **integral or expectation,** $\mathbb{E}(f)=\mathbb{E}_P(f)$, is defined by*
$$\mathbb{E}_P(f)=\int_X f dP=\sum_{y\in\mathbb{C}}yP(f=y). \tag{4.13}$$

Example 4.17. Suppose that $A\in\mathcal{A}$, then
$$\mathbb{E}1_A=0\cdot P(A^c)+1\cdot P(A)=P(A). \tag{4.14}$$

Remark 4.18. Let us recall that our intuitive notion of $P(A)$ was given as in Eq. (2.1) by
$$P(A)=\lim_{N\to\infty}\frac{1}{N}\sum 1_A(\omega(k))$$
where $\omega(k)\in\Omega$ was the result of the k^{th} “independent” experiment. If we use this interpretation back in Eq. (4.13) we arrive at,

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

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

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

1. *If $f\in\mathbb{S}(\mathcal{A})$ and $\lambda\in\mathbb{C}$, then*
$$\mathbb{E}(\lambda f)=\lambda\mathbb{E}(f). \tag{4.15}$$
2. *If $f,g\in\mathbb{S}(\mathcal{A})$, then*
$$\mathbb{E}(f+g)=\mathbb{E}(g)+\mathbb{E}(f). \tag{4.16}$$
   Items 1. and 2. say that $\mathbb{E}(\cdot)$ is a linear functional on $\mathbb{S}(\mathcal{A})$.
3. *If $f=\sum_{j=1}^N\lambda_j 1_{A_j}$ for some $\lambda_j\in\mathbb{C}$ and some $A_j\in\mathbb{C}$, then*
$$\mathbb{E}f=\sum_{j=1}^N\lambda_j P(A_j). \tag{4.17}$$
4. *$\mathbb{E}$ is **positive**, i.e. $\mathbb{E}(f)\geq 0$ for all $0\leq f\in\mathbb{S}(\mathcal{A})$.*
5. *For all $f\in\mathbb{S}(\mathcal{A})$,*
$$|\mathbb{E}f|\leq\mathbb{E}|f|. \tag{4.18}$$

Proof.

1. If $\lambda\neq 0$, then
$$\begin{aligned}\mathbb{E}(\lambda f)&=\sum_{y\in\mathbb{C}}y\ P(\lambda f=y)=\sum_{y\in\mathbb{C}}y\ P(f=y/\lambda)\\ &=\sum_{z\in\mathbb{C}}\lambda z\ P(f=z)=\lambda\mathbb{E}(f).\end{aligned}$$
   The case $\lambda=0$ is trivial.

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

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

But

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

and similarly,

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

Equation (4.16) is now a consequence of the last three displayed equations.

3. If $f = \sum_{j=1}^N \lambda_j 1_{A_j}$, then

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

4. If $f \geq 0$ then

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

5. By the triangle inequality,

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

wherein the last equality we have used Eq. (4.17) and the fact that $|f| = \sum_{\lambda\in\mathbb{C}} |\lambda|\, 1_{f=\lambda}$.

■

Remark 4.20. If Ω is a finite set and $\mathcal{A} = 2^\Omega$, then

$$f(\cdot) = \sum_{\omega\in\Omega} f(\omega)\, 1_{\{\omega\}}$$

and hence

$$\mathbb{E}_P f = \sum_{\omega\in\Omega} f(\omega)\, P(\{\omega\}).$$

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

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

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

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

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

$$P(\{|f| \geq \varepsilon\}) = \mathbb{E}\left[1_{|f|\geq\varepsilon}\right] \leq \mathbb{E}\left[\frac{|f|^p}{\varepsilon^p} 1_{|f|\geq\varepsilon}\right] \leq \varepsilon^{-p}\mathbb{E}|f|^p. \tag{4.19}$$

Observe that

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

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

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

$$\mu\left(\cup_{n=1}^M A_n\right) = \sum_{k=1}^M (-1)^{k+1} \sum_{1\leq n_1<n_2<\cdots<n_k\leq M} \mu(A_{n_1}\cap\cdots\cap A_{n_k}). \tag{4.20}$$

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

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

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

from which it follows that

$$1_{\cup_{n=1}^M A_n} = 1_A = \sum_{k=1}^M (-1)^{k+1} \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} 1_{A_{n_1} \cap \cdots \cap A_{n_k}}. \tag{4.21}$$

Integrating this identity with respect to μ gives Eq. (4.20). ■

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

$$\begin{aligned} \mu\left(\cup_{n=1}^M A_n\right) &+ \sum_{k=2 \ \& \ k \text{ even}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right) \\ &= \sum_{k=1 \ \& \ k \text{ odd}}^M \sum_{1 \le n_1 < n_2 < \cdots < n_k \le M} \mu\left(A_{n_1} \cap \cdots \cap A_{n_k}\right). \end{aligned} \tag{4.22}$$

This can be proved by moving every term with a negative sign on the right side of Eq. (4.21) to the left side and then integrate the resulting identity. Alternatively, Eq. (4.22) follows directly from Eq. (4.20) if $\mu\left(\cup_{n=1}^M A_n\right) < \infty$ and when $\mu\left(\cup_{n=1}^M A_n\right) = \infty$ one easily verifies that both sides of Eq. (4.22) are infinite.

To better understand Eq. (4.21), consider the case $M = 3$ where,

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

so that

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

Here is an alternate proof of Eq. (4.21). Let $\omega \in \Omega$ and by relabeling the sets $\{A_n\}$ if necessary, we may assume that $\omega \in A_1 \cap \cdots \cap A_m$ and $\omega \notin A_{m+1} \cup \cdots \cup A_M$ for some $0 \le m \le M$. (When $m = 0$, both sides of Eq. (4.21) are zero and so we will only consider the case where $1 \le m \le M$.) With this notation we have

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

This verifies Eq. (4.21) since $1_{\cup_{n=1}^M A_n}(\omega) = 1$.

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

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

Since

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

and

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

we find

$$P(B) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n (-1)^{k+1} \frac{1}{k!}. \tag{4.23}$$

For large n this gives,

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

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

$$X_i = 1_{A_i}$$

and observe that

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

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

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

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

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

and so

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

while

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

and

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

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

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

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

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

1. (as functions on Ω) that

$$\binom{N}{k} = \sum_{\Lambda \subset \{1,2,\ldots,n\} \ni |\Lambda| = k} 1_{\cap_{i \in \Lambda} A_i}, \tag{4.24}$$

where by definition

$$\binom{m}{k} = \begin{cases} 0 & \text{if } \quad k > m \\ \frac{m!}{k! \cdot (m-k)!} & \text{if } 1 \le k \le m \\ 1 & \text{if } \quad k = 0 \end{cases}. \tag{4.25}$$

2. Concluded from Eq. (4.24) that for all $z \in \mathbb{C}$,

$$(1+z)^N = 1 + \sum_{k=1}^{n} z^k \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}} \tag{4.26}$$

for all $z \in \mathbb{C}$ provided $(1+z)^0 = 1$ even when $z = -1$.

3. Concluded from Eq. (4.24) to conclude that $S_k = \mathbb{E}_P \binom{N}{k}$.

Exercise 4.5. Taking expectations of Eq. (4.26) implies,

$$\mathbb{E}\left[(1+z)^N\right] = 1 + \sum_{k=1}^{n} S_k z^k. \tag{4.27}$$

Show that setting $z = -1$ in Eq. (4.27) gives another proof of the inclusion exclusion formula. **Hint:** use the definition of the expectation to write out $\mathbb{E}\left[(1+z)^N\right]$ explicitly.

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

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

Hint: differentiate Eq. (4.27) m times with respect to z and then evaluate the result at $z=-1$. In order to do this you will find it useful to derive formulas for;

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

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

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

Therefore it follows from Exercise 4.6 that

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

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

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

Let us check our results are consistent with Eq. (4.23);

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

wherein we have used,

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

4.3.1 Appendix: Bonferroni Inequalities

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

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

$$\sum_{l=0}^{k} (-1)^l \binom{n}{l} = (-1)^k \binom{n-1}{k} 1_{n>0} + 1_{n=0}. \tag{4.28}$$

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

First proof. Just use induction on k. When $k=0$, Eq. (4.28) holds since $1=1$. The induction step is as follows,

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

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

$$\begin{aligned}m_k:=\sum_{l=0}^{k}(-1)^l\binom{n}{l}&=\sum_{l=0}^{k}(-1)^l\cdot\#\left(\Lambda\in 2^X:\#(\Lambda)=l\right)\\&=\sum_{\Lambda\in 2^X:\ \#(\Lambda)\le k}(-1)^{\#(\Lambda)}\end{aligned}\tag{4.29}$$

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

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

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

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

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

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

we see that $\sum_{\Lambda\in\Gamma_k}(-1)^{\#(\Lambda)}=0$. Using this observation with Eq. (4.29) implies

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

■

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

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

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

$$\mu\left(\cup_{n=1}^M A_n\right)=\sum_{l=1}^{k}(-1)^{l+1}S_l+(-1)^k\,\mathbb{E}_\mu\left[\binom{N-1}{k}\right].\tag{4.30}$$

This leads to the Bonferroni inequalities;

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

and

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

Proof. By Lemma 4.28,

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

Therefore integrating this equation with respect to μ gives,

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

and therefore,

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

The Bonferroni inequalities are a simple consequence of Eq. (4.30) and the fact that

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

■

4.3.2 Appendix: Riemann Stieljtes integral

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

$$\mathbb{E}_\mu f:=\int_X f\,d\mu:=\sum_{\lambda\in\mathbb{C}}\lambda\mu(f=\lambda)\ \forall\ f\in\mathbb{S}(\mathcal{A}).\tag{4.31}$$

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

Exercise 4.7. Prove the following statements.

1. For all $f \in \mathbb{S}(\mathcal{A})$,
$$|\mathbb{E}_\mu f| \le \mu(X) \|f\|_\infty . \tag{4.32}$$
2. If $f \in \bar{\mathbb{S}}$ and $f_n \in \mathbb{S} := \mathbb{S}(\mathcal{A})$ such that $\lim_{n\to\infty} \|f - f_n\|_u = 0$, show $\lim_{n\to\infty} \mathbb{E}_\mu f_n$ exists. Also show that defining $\mathbb{E}_\mu f := \lim_{n\to\infty} \mathbb{E}_\mu f_n$ is well defined, i.e. you must show that $\lim_{n\to\infty} \mathbb{E}_\mu f_n = \lim_{n\to\infty} \mathbb{E}_\mu g_n$ if $g_n \in \mathbb{S}$ such that $\lim_{n\to\infty} \|f - g_n\|_u = 0$.
3. Show $\mathbb{E}_\mu : \bar{\mathbb{S}} \to \mathbb{C}$ is still linear and still satisfies Eq. (4.32).
4. Show $|f| \in \bar{\mathbb{S}}$ if $f \in \bar{\mathbb{S}}$ and that Eq. (4.18) is still valid, i.e. $|\mathbb{E}_\mu f| \le \mathbb{E}_\mu |f|$ for all $f \in \bar{\mathbb{S}}$.

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

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

1. *there exists a unique finitely additive measure,* μ_F, *on* $\mathcal{A} := \mathcal{A}(\mathcal{S})$ *such that* $\mu_F((a,b]) = F(b) - F(a)$ *for all* $0 \le a \le b \le T$ *and* $\mu_F(\{0\}) = 0$. *(In fact on could allow for* $\mu_F(\{0\}) = \lambda$ *for any* $\lambda \ge 0$, *but we would then have to write* $\mu_{F,\lambda}$ *rather than* μ_F.*)*
2. *Show* $C([0,1],\mathbb{C}) \subset \overline{\mathbb{S}(\mathcal{A})}$. *More precisely, suppose* $\pi := \{0 = t_0 < t_1 < \dots < t_n = T\}$ *is a partition of* $[0,T]$ *and* $c = (c_1, \dots, c_n) \in [0,T]^n$ *with* $t_{i-1} \le c_i \le t_i$ *for each* i. *Then for* $f \in C([0,1],\mathbb{C})$, *let*
$$f_{\pi,c} := f(0) 1_{\{0\}} + \sum_{i=1}^n f(c_i) 1_{(t_{i-1},t_i]}. \tag{4.33}$$
Show that $\|f - f_{\pi,c}\|_u$ *is small provided,* $|\pi| := \max\{|t_i - t_{i-1}| : i = 1,2,\dots,n\}$ *is small.*
3. *Using the above results, show*
$$\int_{[0,T]} f d\mu_F = \lim_{|\pi|\to 0} \sum_{i=1}^n f(c_i)(F(t_i) - F(t_{i-1}))$$
where the c_i *may be chosen arbitrarily subject to the constraint that* $t_{i-1} \le c_i \le t_i$.

It is customary to write $\int_0^T f dF$ for $\int_{[0,T]} f d\mu_F$. This integral satisfies the estimates,
$$\left| \int_{[0,T]} f d\mu_F \right| \le \int_{[0,T]} |f| \, d\mu_F \le \|f\|_u (F(T) - F(0)) \ \forall\ f \in \overline{\mathbb{S}(\mathcal{A})}.$$
When $F(t) = t$,
$$\int_0^T f dF = \int_0^T f(t) \, dt,$$
is the usual Riemann integral.

Exercise 4.8. Let $a \in (0,T)$, $\lambda > 0$, and
$$G(x) = \lambda \cdot 1_{x\ge a} = \begin{cases} \lambda \text{ if } x \ge a \\ 0 \text{ if } x < a \end{cases}.$$

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

Exercise 4.9. Suppose that $F, G : [0,T] \to \mathbb{R}$ are two increasing functions such that $F(0) = G(0)$, $F(T) = G(T)$, and $F(x) \ne G(x)$ for at most countably many points, $x \in (0,T)$. Show
$$\int_{[0,T]} f d\mu_F = \int_{[0,T]} f d\mu_G \text{ for all } f \in C([0,1],\mathbb{C}). \tag{4.34}$$

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

One of the points of the previous exercise is to show that Eq. (4.34) holds when $G(x) := F(x+)$ – the right continuous version of F. The exercise applies since and increasing function can have at most countably many jumps. So if we only want to integrate continuous functions, we may always assume that $F : [0,T] \to \mathbb{R}$ is right continuous.

4.4 Simple Independence and the Weak Law of Large Numbers

To motivate the exercises in this section, let us imagine that we are following the outcomes of two "independent" experiments with values $\{\alpha_k\}_{k=1}^\infty \subset \Lambda_1$ and

$\{\beta_k\}_{k=1}^{\infty} \subset \Lambda_2$ where Λ_1 and Λ_2 are two finite set of outcomes. Here we are using term independent in an intuitive form to mean that knowing the outcome of one of the experiments gives us no information about outcome of the other.

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

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

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

$$\begin{aligned} p(x,y) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \alpha_k = x \text{ and } \beta_k = y\}, \\ p_1(x) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \alpha_k = x\}, \text{ and} \\ p_2(y) &:= \lim_{N\to\infty} \frac{1}{N} \# \{1 \le k \le N : \beta_k = y\}. \end{aligned} \tag{4.35}$$

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

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

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

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

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

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

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

$$p(x,y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} 1_{\{\beta_k=y\}} \cdot \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} 1_{\left\{\alpha_k^{\ell}=x\right\}}. \tag{4.36}$$

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

$$\lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} 1_{\left\{\alpha_k^{\ell}=x\right\}} = p_1(x) \tag{4.37}$$

independent of the choice of k. Therefore combining Eqs. (4.35), (4.36), and (4.37) implies,

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

■

To understand this "Lemma" in another but equivalent way, let $X_1 : \Lambda_1 \times \Lambda_2 \to \Lambda_1$ and $X_2 : \Lambda_1 \times \Lambda_2 \to \Lambda_2$ be the projection maps, $X_1(x,y) = x$ and $X_2(x,y) = y$ respectively. Further suppose that $f : \Lambda_1 \to \mathbb{R}$ and $g : \Lambda_2 \to \mathbb{R}$ are functions, then using the heuristics Lemma 4.32 implies,

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

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

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

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

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

$$P(A) := \sum_{\omega\in A} p(\omega) \text{ for all } A \in 2^{\Omega}. \tag{4.38}$$

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

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

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

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

$$\mathbb{E}_P\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}_P[f_i(X_i)] \tag{4.39}$$

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

Definition 4.33 (Independence). *We say simple random variables, $X_1,\dots,X_n$ with values in Λ on some probability space, $(\Omega,\mathcal{A},P)$ are independent (more precisely P – independent) if Eq. (4.39) holds for all functions, $f_i : \Lambda \to \mathbb{R}$.*

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

$$P(X_1 \in A_1,\dots,X_n \in A_n) = P(X_1 \in A_1)\dots P(X_n \in A_n) \tag{4.40}$$

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

Exercise 4.13 (A Weak Law of Large Numbers). Suppose that $\Lambda \subset \mathbb{R}$ is a finite set, $n \in \mathbb{N}$, $\Omega = \Lambda^n$, $p(\omega) = \prod_{i=1}^n q(\omega_i)$ where $q : \Lambda \to [0,1]$ such that $\sum_{\lambda\in\Lambda} q(\lambda) = 1$, and let $P : 2^\Omega \to [0,1]$ be the probability measure defined as in Eq. (4.38). Further let $X_i(\omega) = \omega_i$ for $i = 1,2,\dots,n$, $\xi := \mathbb{E}X_i$, $\sigma^2 := \mathbb{E}(X_i - \xi)^2$, and

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

1. Show, $\xi = \sum_{\lambda\in\Lambda} \lambda\, q(\lambda)$ and
$$\sigma^2 = \sum_{\lambda\in\Lambda} (\lambda - \xi)^2 q(\lambda) = \sum_{\lambda\in\Lambda} \lambda^2 q(\lambda) - \xi^2. \tag{4.41}$$
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
$$\mathbb{E}(S_n - \xi)^2 = \frac{1}{n}\sigma^2. \tag{4.42}$$
5. Conclude using Eq. (4.42) and Remark 4.22 that
$$P(|S_n - \xi| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2. \tag{4.43}$$

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

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

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

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

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

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

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

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

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

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

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

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

Then

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

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

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

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

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

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

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

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

Hence we find

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

where

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

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

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

and hence we may conclude that

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

and therefore, that

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

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

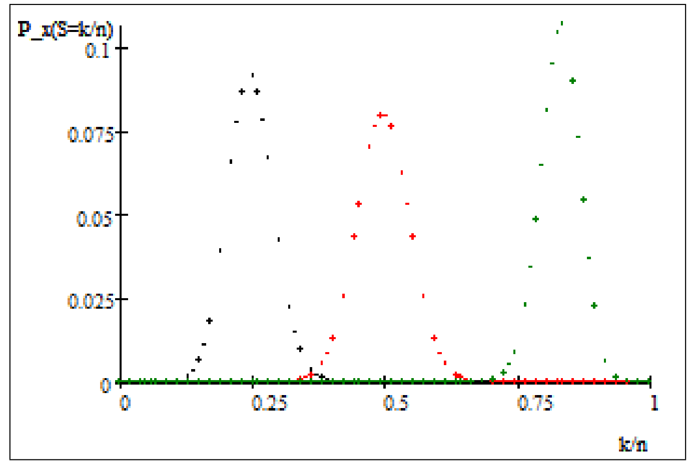


Fig. 4.1. Plots of $P_x(S_n = k/n)$ versus k/n for $n = 100$ with $x = 1/4$ (black), $x = 1/2$ (red), and $x = 5/6$ (green).

■

4.4.1 Product Measures and Fubini's Theorem

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

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

$$C = \sum_{i=1}^{n} A_i \times B_i \text{ with } A_i \times B_i \in \mathcal{S}. \tag{4.44}$$

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

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

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

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

Proof. According to Eq. (4.44),

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

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

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

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

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

Similarly one shows that

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

In particular this shows that we may define

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

and with this definition we have,

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

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

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

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

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

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

4.5 Simple Conditional Expectation

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

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

$$\bar{X}(\omega) := \frac{1}{P(A_\omega)}\mathbb{E}[1_{A_\omega}X] \text{ for all } \omega \in \Omega, \quad (4.45)$$

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

$$\bar{X} = \sum_{i=1}^{n} \frac{\mathbb{E}[1_{B_i}X]}{P(B_i)} 1_{B_i}, \quad (4.46)$$

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

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

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

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

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

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

1. **(Orthogonal Projection Property 1.)** If $Z \in \mathbb{S}(\mathcal{A})$, then

$$\mathbb{E}[XZ] = \mathbb{E}[\bar{X}Z] = \mathbb{E}[\mathbb{E}_{\mathcal{A}}X \cdot Z] \quad (4.47)$$

and

$$(\mathbb{E}_{\mathcal{A}}Z)(\omega) = \begin{cases} Z(\omega) & \text{if } P(A_\omega) > 0 \\ 0 & \text{if } P(A_\omega) = 0 \end{cases}. \quad (4.48)$$

In particular, $\mathbb{E}_{\mathcal{A}}[\mathbb{E}_{\mathcal{A}}Z] = \mathbb{E}_{\mathcal{A}}Z$.
This basically says that $\mathbb{E}_{\mathcal{A}}$ is orthogonal projection from $\mathbb{S}(\mathcal{B})$ onto $\mathbb{S}(\mathcal{A})$ relative to the inner product

$$(f, g) = \mathbb{E}[fg] \text{ for all } f, g \in \mathbb{S}(\mathcal{B}).$$

2. **(Orthogonal Projection Property 2.)** If $Y \in \mathbb{S}(\mathcal{A})$ satisfies, $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all $Z \in \mathbb{S}(\mathcal{A})$, then $Y(\omega) = \bar{X}(\omega)$ whenever $P(A_\omega) > 0$. In particular, $P(Y \neq \bar{X}) = 0$. **Hint:** use item 1. to compute $\mathbb{E}\left[(\bar{X} - Y)^2\right]$.
3. **(Best Approximation Property.)** For any $Y \in \mathbb{S}(\mathcal{A})$,

$$\mathbb{E}\left[(X - \bar{X})^2\right] \le \mathbb{E}\left[(X - Y)^2\right] \quad (4.49)$$

with equality iff $\bar{X} = Y$ almost surely (a.s. for short), where $\bar{X} = Y$ a.s. iff $P(\bar{X} \neq Y) = 0$. In words, $\bar{X} = \mathbb{E}_{\mathcal{A}}X$ is the best ("L^2") approximation to X by an $\mathcal{A}$ – measurable random variable.
4. **(Contraction Property.)** $\mathbb{E}|\bar{X}| \le \mathbb{E}|X|$. (It is typically **not** true that $|\bar{X}(\omega)| \le |X(\omega)|$ for all ω.)
5. (**Pull Out Property.**) If $Z \in \mathbb{S}(\mathcal{A})$, then

$$\mathbb{E}_{\mathcal{A}}[ZX] = Z\mathbb{E}_{\mathcal{A}}X.$$

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

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

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

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

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

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

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

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

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

$$\mathbb{E}_{\mathcal{A}}[g(Y)](x,y) = \mathbb{E}[g(Y)] \text{ for all } (x,y) \in \Omega. \tag{4.50}$$

Let us further observe that if $f : \Lambda_1 \to \mathbb{R}$ is any other function, then $f(X)$ is an $\mathcal{A}$ – simple function and therefore by Eq. (4.50) and Exercise 4.15

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

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

Lemma 4.42 (Conditional Expectation and Independence). *Let $\Omega = \Lambda_1 \times \Lambda_2$, X, Y, $\mathcal{B} = 2^{\Omega}$, and $\mathcal{A} = X^{-1}\left(2^{\Lambda_1}\right)$, be as in Example 4.41 above. Assume that $P : \mathcal{B} \to [0,1]$ is a probability measure. If X and Y are P – independent, then Eq. (4.50) holds.*

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

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

■

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

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

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

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

Countably Additive Measures

Proposition 5.1. *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 $A_n \downarrow \emptyset$, $P(A_n) \downarrow 0$.*

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

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

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

$2 \implies 1$. 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 5.2. 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 5.3. *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 5.1 hold. We will call $(\Omega, \mathcal{B}, P)$ a probability space.*

Lemma 5.4. *Suppose that $(\Omega, \mathcal{B}, P)$ is a probability space, then P is countably sub-additive.*

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

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

■

Example 5.5 (Outline of Construction of Measures on $\mathbb{R}$). Suppose that $\Omega = \mathbb{R}$, $\mathcal{S} := \{(a,b] \cap \mathbb{R} : -\infty \leq a \leq b \leq \infty\}$, and $\mathcal{A} = \mathcal{A}(\mathcal{S})$ consists of those sets, $A \subset \mathbb{R}$ which may be written as finite disjoint unions of sets from $\mathcal{S}$ as in Example 3.26. Recall that $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A}) = \sigma(\mathcal{S})$.

1. We have seen in Proposition 4.8 that to every increasing function, $F : \mathbb{R} \to [0,1]$ such that

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

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

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

2. We then show in Lemma 5.8 and Proposition 5.12 that P is σ – additive on $\mathcal{A}$ iff F is right continuous.
3. We will then appeal to Theorem 5.25 to see that P extends to a probability measure on $\mathcal{B}_{\mathbb{R}}$ iff F is right continuous.

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

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

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

Example 5.7. Suppose that

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

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

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

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

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

Proof. The monotonicity of P shows that $F(x)$ in Eq. (5.1) is non-decreasing. For $a \in \mathbb{R}$, we

$$\begin{aligned} F(a + 1/n) - F(a) &= P((-\infty, a + 1/n]) - P((-\infty, a]) \\ &= P((a, a + 1/n]) \downarrow 0 \end{aligned}$$

by the continuity of P from above. Therefore, $F(a+) := \lim_{x \downarrow a} F(x) = F(a)$ for all $a \in \mathbb{R}$ and hence F is right continuous. Similar arguments show that $F(\infty) = 1$ and $F(-\infty) = 0$. ■

Theorem 5.9. *To each function $F : \mathbb{R} \to [0,1]$ satisfying properties 1. – 4. in Lemma 5.8, there exists a unique probability measure, P_F, on $\mathcal{B}_{\mathbb{R}}$ such that*

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

Proof. The uniqueness assertion in the theorem is covered in Exercise 5.1 below. The existence portion of the Theorem follows from Proposition 5.12 and Theorem 5.25 below. ■

Example 5.10 (Uniform Distribution). The function,

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

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 $(0,1]$. $\min(1, \max(0, x))$

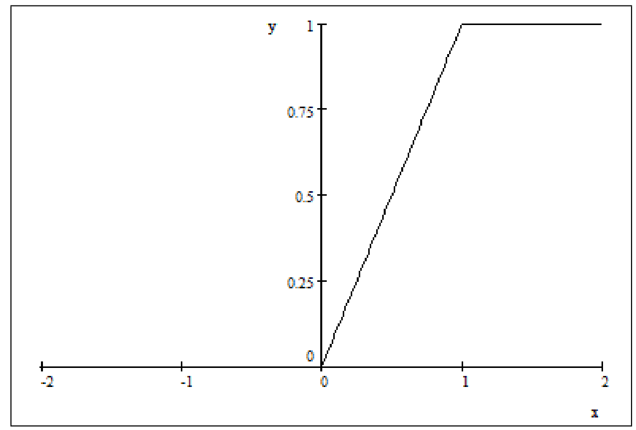


Fig. 5.1. The cumulative distribution function for the uniform distribution.

Recall from Definition 3.12 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 5.11. *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,

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

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

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

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

Now suppose that $F : \mathbb{R} \to \mathbb{R}$ is an increasing function, $F(\pm\infty) := \lim_{x \to \pm\infty} F(x)$, and $\mu = \mu_F$ is the finitely additive measure on $(\mathbb{R}, \mathcal{A})$ described in Proposition 4.8. If μ happens to be a premeasure on $\mathcal{A}$, then we have seen that F must be right continuous.[1] 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 modulo translation by constants. The latter ambiguity may be fixed by requiring $F(0) = 0$.

Proposition 5.12. *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*

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

[1] For sake of completeness here is the argument again. 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 $\lim_{y \downarrow b} F(y) = F(b)$.

Proof. As above, let $F(\pm\infty) := \lim_{x \to \pm\infty} F(x)$ and $\mu = \mu_F$ be as in Remark 4.9. Because of Proposition 5.11, to finish the proof it suffices to show μ is sub-additive on $\mathcal{S}$. Thus we must show

$$\mu(J) \le \sum_{n=1}^{\infty} \mu(J_n). \tag{5.2}$$

where $J = \sum_{n=1}^{\infty} J_n$ with $J = (a,b] \cap \mathbb{R}$ and $J_n = (a_n, b_n] \cap \mathbb{R}$. Let us recall from Proposition 4.2 that the finite additivity of μ implies

$$\sum_{n=1}^{\infty} \mu(J_n) \le \mu(J). \tag{5.3}$$

We begin with the special case where $-\infty < a < b < \infty$. Our proof will be by "continuous induction." The strategy is to show $a \in \Lambda$ where

$$\Lambda := \left\{ \alpha \in [a,b] : \mu(J \cap (\alpha, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha, b]) \right\}. \tag{5.4}$$

As $b \in J$, there exists an k such that $b \in J_k$ and hence $(a_k, b_k] = (a_k, b]$ for this k. It now easily follows that $J_k \subset \Lambda$ so that Λ is not empty. To finish the proof we are going to show $\bar{a} := \inf \Lambda \in \Lambda$ and that $\bar{a} = a$.

- If $\bar{a} \notin \Lambda$, there would exist $\alpha_m \in \Lambda$ such that $\alpha_m \downarrow \bar{a}$, i.e.

$$\mu(J \cap (\alpha_m, b]) \le \sum_{n=1}^{\infty} \mu(J_n \cap (\alpha_m, b]). \tag{5.5}$$

  Since $\mu(J_n \cap (\alpha_m, b]) \le \mu(J_n)$ and $\sum_{n=1}^{\infty} \mu(J_n) \le \mu(J) < \infty$ by Eq. (5.3), we may use the right continuity of F and the dominated convergence theorem for sums in order to pass to the limit as $m \to \infty$ in Eq. (5.5) to learn,

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

  This shows $\bar{a} \in \Lambda$ which is a contradiction to the original assumption that $\bar{a} \notin \Lambda$.
- If $\bar{a} > a$, then $\bar{a} \in J_l = (a_l, b_l]$ for some l. Letting $\alpha = a_l < \bar{a}$, we have,

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

This shows $\alpha\in\Lambda$ and $\alpha<\bar{a}$ which violates the definition of $\bar{a}$. Thus we must conclude that $\bar{a}=a$.

The hard work is now done but we still have to check the cases where $a=-\infty$ or $b=\infty$. For example, suppose that $b=\infty$ so that

$$J=(a,\infty)=\sum_{n=1}^{\infty}J_n$$

with $J_n=(a_n,b_n]\cap\mathbb{R}$. Then

$$I_M:=(a,M]=J\cap I_M=\sum_{n=1}^{\infty}J_n\cap I_M$$

and so by what we have already proved,

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

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

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

The other cases where $a=-\infty$ and $b\in\mathbb{R}$ and $a=-\infty$ and $b=\infty$ are handled similarly. ∎

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*

Technically we only need Definition 5.13 and Lemma 5.14 from this section.

Definition 5.13. *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 5.14. *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 5.15 (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 $i=1,2,\dots$ and $\varepsilon>0$ is given. By assumption there exists $A_i\in\mathcal{A}_\delta$ and $C_i\in\mathcal{A}_\sigma$ such that $A_i\subset B_i\subset C_i$ and $\mu(C_i\setminus A_i)<2^{-i}\varepsilon$.

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

$$C\setminus A=[\cup_{i=1}^\infty C_i]\setminus A=\cup_{i=1}^\infty[C_i\setminus A]\subset\cup_{i=1}^\infty[C_i\setminus A_i].$$

Therefore,

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

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

Many theorems in the sequel will require some control on the size of a measure μ. The relevant notion for our purposes (and most purposes) is that of a σ – finite measure defined next.

Definition 5.16. *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 5.17 (σ – Finite Regularity Result). *Theorem 5.15 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 5.15, 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

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

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 5.18. *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 5.15, 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

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

■

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 5.15.
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 5.19. Let us recall from Proposition 5.1 and Remark 5.2 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 5.20. *Given a premeasure, $\mu : \mathcal{A} \to [0,\infty]$, we extend μ to $\mathcal{A}_\sigma$ by defining*

$$\mu(B) := \sup\{\mu(A) : \mathcal{A} \ni A \subset B\}. \tag{5.6}$$

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

1. *(**Monotonicity**) If $A, B \in \mathcal{A}_\sigma$ with $A \subset B$ then $\mu(A) \le \mu(B)$.*

2. ***(Continuity)*** *If $A_n \in \mathcal{A}$ and $A_n \uparrow A \in \mathcal{A}_\sigma$, then $\mu(A_n) \uparrow \mu(A)$ as $n \to \infty$.*
3. ***(Strong Additivity)*** *If $A, B \in \mathcal{A}_\sigma$, then*
$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B). \tag{5.7}$$
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*
$$\mu(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.8}$$
5. *(σ - Additivity **on** $\mathcal{A}_\sigma$) The function μ is countably additive on $\mathcal{A}_\sigma$.*

Proof. 1. and 2. Monotonicity follows directly from Eq. (5.6) which then implies $\mu(A_n) \le \mu(B)$ for all n. Therefore $M := \lim_{n\to\infty} \mu(A_n) \le \mu(B)$. To prove the reverse inequality, let $\mathcal{A} \ni A \subset B$. Then by the continuity of μ on $\mathcal{A}$ and the fact that $A_n \cap A \uparrow A$ we have $\mu(A_n \cap A) \uparrow \mu(A)$. As $\mu(A_n) \ge \mu(A_n \cap A)$ for all n it then follows that $M := \lim_{n\to\infty} \mu(A_n) \ge \mu(A)$. As $A \in \mathcal{A}$ with $A \subset B$ was arbitrary we may conclude,
$$\mu(B) = \sup\{\mu(A) : \mathcal{A} \ni A \subset B\} \le M.$$

3. Suppose that $A, B \in \mathcal{A}_\sigma$ and $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ are sequences in $\mathcal{A}$ such that $A_n \uparrow A$ and $B_n \uparrow B$ as $n \to \infty$. Then passing to the limit as $n \to \infty$ in the identity,
$$\mu(A_n \cup B_n) + \mu(A_n \cap B_n) = \mu(A_n) + \mu(B_n)$$
proves Eq. (5.7). In particular, it follows that μ is finitely additive on $\mathcal{A}_\sigma$.

4 and 5. Let $\{A_n\}_{n=1}^\infty$ be any sequence in $\mathcal{A}_\sigma$ and choose $\{A_{n,i}\}_{i=1}^\infty \subset \mathcal{A}$ such that $A_{n,i} \uparrow A_n$ as $i \to \infty$. Then we have,
$$\mu\left(\cup_{n=1}^N A_{n,N}\right) \le \sum_{n=1}^N \mu(A_{n,N}) \le \sum_{n=1}^N \mu(A_n) \le \sum_{n=1}^\infty \mu(A_n). \tag{5.9}$$
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. (5.9) to conclude Eq. (5.8) holds. If we further assume that $\{A_n\}_{n=1}^\infty \subset \mathcal{A}_\sigma$ are pairwise disjoint, by the finite additivity and monotonicity of μ on $\mathcal{A}_\sigma$, we have
$$\sum_{n=1}^\infty \mu(A_n) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n) = \lim_{N\to\infty} \mu\left(\cup_{n=1}^N A_n\right) \le \mu(\cup_{n=1}^\infty A_n).$$
This inequality along with Eq. (5.8) shows that μ is $\sigma-$ 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
$$\mu(A) := \mu(X) - \mu(A^c) \text{ for all } A \in \mathcal{A}_\delta. \tag{5.10}$$

Lemma 5.21. *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 5.20 and Eq. (5.10) above.*

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

Proof.

1. Since $\mu(B) = \mu(X) - \mu(B^c)$ and $A \subset B$ iff $B^c \subset A^c$, it follows that
$$\begin{aligned}\inf\{\mu(B) : A \subset B \in \mathcal{A}\} &= \inf\{\mu(X) - \mu(B^c) : \mathcal{A} \ni B^c \subset A^c\} \\ &= \mu(X) - \sup\{\mu(B) : \mathcal{A} \ni B \subset A^c\} \\ &= \mu(X) - \mu(A^c) = \mu(A).\end{aligned}$$
2. Similarly, since $A_n^c \uparrow A^c \in \mathcal{A}_\sigma$, by the definition of $\mu(A)$ and Proposition 5.20 it follows that
$$\begin{aligned}\mu(A) = \mu(X) - \mu(A^c) &= \mu(X) - \uparrow \lim_{n\to\infty} \mu(A_n^c) \\ &= \downarrow \lim_{n\to\infty} [\mu(X) - \mu(A_n^c)] = \downarrow \lim_{n\to\infty} \mu(A_n).\end{aligned}$$
3. Suppose $A, B \in \mathcal{A}_\delta$ and $A_n, B_n \in \mathcal{A}$ such that $A_n \downarrow A$ and $B_n \downarrow B$, then $A_n \cup B_n \downarrow A \cup B$ and $A_n \cap B_n \downarrow A \cap B$ and therefore,
$$\begin{aligned}\mu(A \cup B) + \mu(A \cap B) &= \lim_{n\to\infty} [\mu(A_n \cup B_n) + \mu(A_n \cap B_n)] \\ &= \lim_{n\to\infty} [\mu(A_n) + \mu(B_n)] = \mu(A) + \mu(B).\end{aligned}$$
4. Since $A^c, C \in \mathcal{A}_\sigma$ we may use the strong additivity of μ on $\mathcal{A}_\sigma$ to conclude,
$$\mu(A^c \cup C) + \mu(A^c \cap C) = \mu(A^c) + \mu(C).$$
Because $X = A^c \cup C$, and $\mu(A^c) = \mu(X) - \mu(A)$, the above equation may be written as
$$\mu(X) + \mu(C \setminus A) = \mu(X) - \mu(A) + \mu(C)$$
which finishes the proof.

■

Notation 5.22 (Inner and outer measures) *Let $\mu : \mathcal{A} \to [0,\infty)$ be a finite premeasure extended to $\mathcal{A}_\sigma \cup \mathcal{A}_\delta$ as above. The for **any** $B \subset X$ let*

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

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

If $B \subset X$ has the same inner and outer content it is reasonable to define the measure of B as this common value. As we will see in Theorem 5.25 below, this extension becomes a σ – additive measure on a σ – algebra of subsets of X.

Definition 5.23 (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 $\mu_*(B) = \mu^*(B)$. We will denote the collection of measurable subsets of X by $\mathcal{B} = \mathcal{B}(\mu)$ and define $\bar{\mu} : \mathcal{B} \to [0, \mu(X)]$ by*

$$\bar{\mu}(B) := \mu_*(B) = \mu^*(B) \text{ for all } B \in \mathcal{B}. \tag{5.11}$$

Remark 5.24. Observe that $\mu_*(B) = \mu^*(B)$ iff for all $\varepsilon > 0$ there exists $A \in \mathcal{A}_\delta$ and $C \in \mathcal{A}_\sigma$ such that $A \subset B \subset C$ such that

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

wherein we have used Lemma 5.21 for the first equality. Moreover we will use below for any $\mathcal{A}_\delta \ni A \subset B \subset C \in \mathcal{A}_\sigma$ that

$$\mu(A) \le \mu_*(B) = \bar{\mu}(B) = \mu^*(B) \le \mu(C), \tag{5.12}$$
$$0 \le \bar{\mu}(B) - \mu(A) \le \mu(C) - \mu(A) = (C \setminus A), \text{ and} \tag{5.13}$$
$$0 \le \mu(C) - \bar{\mu}(B) \le \mu(C) - \mu(A) = \mu(C \setminus A). \tag{5.14}$$

Theorem 5.25 (Finite Premeasure Extension Theorem). *Suppose μ is a finite premeasure on an algebra $\mathcal{A} \subset 2^X$ and $\bar{\mu} : \mathcal{B} := \mathcal{B}(\mu) \to [0, \mu(X)]$ be as in Definition 5.23. 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 5.21 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

$$\sum_{i=1}^\infty \mu(C_i) \le \mu(X) + \varepsilon < \infty. \tag{5.15}$$

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

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

Therefore, using the sub-additivity of μ on $\mathcal{A}_\sigma$ and the estimate (5.15),

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

Since $\varepsilon > 0$ is arbitrary, it follows that $B \in \mathcal{B}$. Moreover by repeated use of Remark 5.24, we find

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

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 $\mathcal{B}$.

You are asked to prove the uniqueness assertion in Exercise 5.3 below. ■

Exercise 5.3. Let μ, $\bar{\mu}$, $\mathcal{A}$, and $\mathcal{B}:=\mathcal{B}(\mu)$ be as in Theorem 5.25. Further suppose that $\mathcal{B}_0\subset 2^X$ is a σ – algebra such that $\mathcal{A}\subset\mathcal{B}_0\subset\mathcal{B}$ and $\nu:\mathcal{B}_0\to[0,\mu(X)]$ is a σ – additive measure on $\mathcal{B}_0$ such that $\nu=\mu$ on $\mathcal{A}$. Show that $v=\bar{\mu}$ on $\mathcal{B}_0$ as well.

Corollary 5.26. *Suppose that $\mathcal{A}\subset 2^X$ is an algebra and $\mu:\mathcal{B}_0:=\sigma(\mathcal{A})\to[0,\mu(X)]$ is a σ – additive measure. Then for every $B\in\sigma(\mathcal{A})$ and $\varepsilon>0$;*

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

Exercise 5.4. Prove corollary 5.26 by considering $\bar{\nu}$ where $\nu:=\mu|_{\mathcal{A}}$. **Hint:** you may find Exercise 4.3 useful here.

Theorem 5.27. *Suppose that μ is a σ – finite premeasure on an algebra $\mathcal{A}$. Then*

$$\bar{\mu}(B):=\inf\{\mu(C):B\subset C\in\mathcal{A}_\sigma\}\ \forall\ B\in\sigma(\mathcal{A}) \tag{5.16}$$

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 5.25 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. (5.16) holds. Let $B\in\sigma(\mathcal{A})$, $B_m=X_m\cap B$ and $\varepsilon>0$ be given. By Theorem 5.25, 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. (5.16). The uniqueness of the extension $\bar{\mu}$ is proved in Exercise 5.1. ■

The following slight reformulation of Theorem 5.27 can be useful.

Corollary 5.28. *Let $\mathcal{A}$ be an algebra of sets, $\{X_m\}_{m=1}^{\infty}\subset\mathcal{A}$ is a given sequence of sets such that $X_m\uparrow X$ as $m\to\infty$. Let*

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

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

Proof. Existence. By assumption, $\mu_m:=\mu|_{\mathcal{A}_{X_m}}:\mathcal{A}_{X_m}\to[0,\infty)$ is a premeasure on $(X_m,\mathcal{A}_{X_m})$ and hence by Theorem 5.27 extends to a measure μ'_m on $(X_m,\sigma(\mathcal{A}_{X_m})=\mathcal{B}_{X_m})$. Let $\bar{\mu}_m(B):=\mu'_m(B\cap X_m)$ for all $B\in\mathcal{B}$. Then $\{\bar{\mu}_m\}_{m=1}^{\infty}$ is an increasing sequence of measure on $(X,\mathcal{B})$ and hence $\bar{\mu}:=\lim_{m\to\infty}\bar{\mu}_m$ defines a measure on $(X,\mathcal{B})$ such that $\bar{\mu}|_{\mathcal{A}_f}=\mu$.

Uniqueness. If μ_1 and μ_2 are two such extensions, then $\mu_1(X_m\cap B)=\mu_2(X_m\cap B)$ for all $B\in\mathcal{A}$ and therefore by Exercise 5.1 or Dynkin's $\pi-\lambda$ theorem below we know that $\mu_1(X_m\cap B)=\mu_2(X_m\cap B)$ for all $B\in\mathcal{B}$. We may now let $m\to\infty$ to see that in fact $\mu_1(B)=\mu_2(B)$ for all $B\in\mathcal{B}$, i.e. $\mu_1=\mu_2$. ■

5.5 Two Important Examples

5.5.1 Lebesgue Measure

If $F(x)=x$ for all $x\in\mathbb{R}$, we denote μ_F by m and call m Lebesgue measure on $(\mathbb{R},\mathcal{B}_\mathbb{R})$.

Theorem 5.29. *Lebesgue measure m is invariant under translations, i.e. for $B\in\mathcal{B}_\mathbb{R}$ and $x\in\mathbb{R}$,*

$$m(x+B)=m(B). \tag{5.17}$$

Lebesgue measure, m, is the unique measure on $\mathcal{B}_{\mathbb{R}}$ such that $m((0,1]) = 1$ and Eq. (5.17) holds for $B \in \mathcal{B}_{\mathbb{R}}$ and $x \in \mathbb{R}$. Moreover, m has the scaling property

$$m(\lambda B) = |\lambda| \, m(B) \tag{5.18}$$

where $\lambda \in \mathbb{R}$, $B \in \mathcal{B}_{\mathbb{R}}$ and $\lambda B := \{\lambda x : x \in B\}$.

Proof. Let $m_x(B) := m(x+B)$, then one easily shows that m_x is a measure on $\mathcal{B}_{\mathbb{R}}$ such that $m_x((a,b]) = b-a$ for all $a < b$. Therefore, $m_x = m$ by the uniqueness assertion in Exercise 5.1. For the converse, suppose that m is translation invariant and $m((0,1]) = 1$. Given $n \in \mathbb{N}$, we have

$$(0,1] = \cup_{k=1}^n \left(\frac{k-1}{n}, \frac{k}{n}\right] = \cup_{k=1}^n \left(\frac{k-1}{n} + (0, \frac{1}{n}]\right).$$

Therefore,

$$\begin{aligned} 1 = m((0,1]) &= \sum_{k=1}^n m\left(\frac{k-1}{n} + (0, \frac{1}{n}]\right) \\ &= \sum_{k=1}^n m((0, \frac{1}{n}]) = n \cdot m((0, \frac{1}{n}]). \end{aligned}$$

That is to say

$$m((0, \frac{1}{n}]) = 1/n.$$

Similarly, $m((0, \frac{l}{n}]) = l/n$ for all $l, n \in \mathbb{N}$ and therefore by the translation invariance of m,

$$m((a,b]) = b - a \text{ for all } a, b \in \mathbb{Q} \text{ with } a < b.$$

Finally for $a, b \in \mathbb{R}$ such that $a < b$, choose $a_n, b_n \in \mathbb{Q}$ such that $b_n \downarrow b$ and $a_n \uparrow a$, then $(a_n, b_n] \downarrow (a,b]$ and thus

$$m((a,b]) = \lim_{n\to\infty} m((a_n, b_n]) = \lim_{n\to\infty} (b_n - a_n) = b - a,$$

i.e. m is Lebesgue measure. To prove Eq. (5.18) we may assume that $\lambda \neq 0$ since this case is trivial to prove. Now let $m_\lambda(B) := |\lambda|^{-1} m(\lambda B)$. It is easily checked that m_λ is again a measure on $\mathcal{B}_{\mathbb{R}}$ which satisfies

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

if $\lambda > 0$ and

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

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

5.5.2 A Baby Version of Kolmogorov's Extension Theorem

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

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

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

Exercise 5.5. Show $\mathcal{A}$ is an algebra.

Lemma 5.30. *Suppose $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ is a decreasing sequence of non-empty cylinder sets, then $\cap_{n=1}^\infty A_n \neq \emptyset$.*

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

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

$$(\omega_1(n), \omega_2(n), \dots, \omega_{N_k}(n)) \in B_k \text{ for all } k \leq n. \tag{5.19}$$

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 $\lambda := (\lambda_1, \lambda_2, \dots) \in \cap_{n=1}^\infty A_n$. By the construction above, for all $N \in \mathbb{N}$ we have

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

Taking $N = N_k$ and $n \in \Gamma_{N_k}$ with $n \geq k$, we learn from Eq. (5.19) 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 5.31 (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 5.25, 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 5.30, 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

$$p_n(\lambda_1, \ldots, \lambda_n) := P(\{\omega \in \Omega : \omega_1 = \lambda_1, \ldots, \omega_n = \lambda_n\}). \tag{5.20}$$

Exercise 5.6 (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 5.7 (Converse to 5.6). 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 5.6 hold. Then there exists a unique probability measure, P on $\sigma(\mathcal{A})$ such that Eq. (5.20) holds for all $n \in \mathbb{N}$ and $(\lambda_1, \ldots, \lambda_n) \in \Lambda^n$.

Example 5.32 (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 5.7 with $p_n(\lambda_1, \ldots, \lambda_n) := q(\lambda_1) \ldots q(\lambda_n)$.

5.6 Completions of Measure Spaces*

Definition 5.33. *A set $E \subset X$ is a **null** set if $E \in \mathcal{B}$ and $\mu(E) = 0$. If P is some "property" which is either true or false for each $x \in X$, we will use the terminology P a.e. (to be read P almost everywhere) to mean*

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

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

Definition 5.34. *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 5.35 (Completion of a Measure). *Let $(X, \mathcal{B}, \mu)$ be a measure space. Set*

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

*see Fig. 5.2. Then $\bar{\mathcal{B}}$ is a σ – algebra, $\bar{\mu}$ is a well defined measure on $\bar{\mathcal{B}}$, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{B}}$ which extends μ on $\mathcal{B}$, and $(X, \bar{\mathcal{B}}, \bar{\mu})$ 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

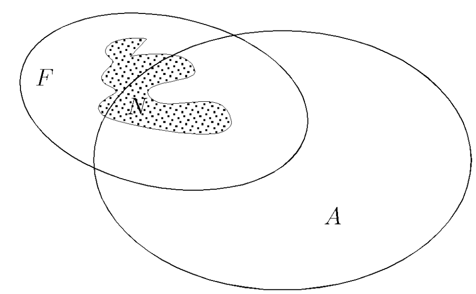


Fig. 5.2. Completing a σ – algebra.

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. ∎

Random Variables

Notation 6.1 *If $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ let*

$$f^{-1}\mathcal{E} := f^{-1}(\mathcal{E}) := \{f^{-1}(E)|E \in \mathcal{E}\}.$$

If $\mathcal{G} \subset 2^X$, let

$$f_*\mathcal{G} := \{A \in 2^Y | f^{-1}(A) \in \mathcal{G}\}.$$

Definition 6.2. *Let $\mathcal{E} \subset 2^X$ be a collection of sets, $A \subset X$, $i_A : A \to X$ be the* ***inclusion map*** *($i_A(x) = x$ for all $x \in A$) and*

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

The following results will be used frequently (often without further reference) in the sequel.

Exercise 6.1. 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 6.1) are algebras (σ – algebras) provided $\mathcal{F}$ and $\mathcal{B}$ are algebras (σ – algebras).

Lemma 6.3. *Suppose that $f : X \to Y$ is a function and $\mathcal{E} \subset 2^Y$ and $A \subset Y$ then*

$$\sigma\left(f^{-1}(\mathcal{E})\right) = f^{-1}(\sigma(\mathcal{E})) \text{ and} \tag{6.1}$$

$$(\sigma(\mathcal{E}))_A = \sigma(\mathcal{E}_A), \tag{6.2}$$

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 6.1, $f^{-1}(\sigma(\mathcal{E}))$ 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. (6.1) 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 6.4. Let $\mathcal{E} = \{(a, b] : -\infty < a < b < \infty\}$ and $\mathcal{B} = \sigma(\mathcal{E})$ be the Borel σ – field on $\mathbb{R}$. Then

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

and we have

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

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

6.1 Measurable Functions

Definition 6.5. *A* ***measurable space*** *is a pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ – algebra on X.*

To motivate the notion of a measurable function, suppose $(X, \mathcal{M}, \mu)$ is a measure space and $f : X \to \mathbb{R}_+$ is a function. Roughly speaking, we are going to define $\int_X f d\mu$ as a certain limit of sums of the form,

$$\sum_{0<a_1<a_2<a_3<\dots}^{\infty} a_i \mu(f^{-1}(a_i, a_{i+1}]).$$

For this to make sense we will need to require $f^{-1}((a, b]) \in \mathcal{M}$ for all $a < b$. Because of Corollary 6.11 below, this last condition is equivalent to the condition $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$.

Definition 6.6. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces. A function $f : X \to Y$ is* ***measurable*** *of more precisely, $\mathcal{M}/\mathcal{F}$ – measurable or $(\mathcal{M}, \mathcal{F})$ – measurable, if $f^{-1}(\mathcal{F}) \subset \mathcal{M}$, i.e. if $f^{-1}(A) \in \mathcal{M}$ for all $A \in \mathcal{F}$.*

Remark 6.7. Let $f : X \to Y$ be a function. Given a σ – algebra $\mathcal{F} \subset 2^Y$, the σ – algebra $\mathcal{M} := f^{-1}(\mathcal{F})$ is the smallest σ – algebra on X such that f is $(\mathcal{M}, \mathcal{F})$ - measurable . Similarly, if $\mathcal{M}$ is a σ - algebra on X then

$$\mathcal{F} = f_*\mathcal{M} = \{A \in 2^Y | f^{-1}(A) \in \mathcal{M}\}$$

is the largest σ – algebra on Y such that f is $(\mathcal{M}, \mathcal{F})$ - measurable.

Example 6.8 (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 6.9. 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 6.10. *Suppose that $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ are measurable spaces and further assume $\mathcal{E} \subset \mathcal{F}$ generates $\mathcal{F}$, i.e. $\mathcal{F} = \sigma(\mathcal{E})$. Then a map, $f : X \to Y$ is measurable iff $f^{-1}(\mathcal{E}) \subset \mathcal{M}$.*

Proof. If f is $\mathcal{M}/\mathcal{F}$ measurable, then $f^{-1}(\mathcal{E}) \subset f^{-1}(\mathcal{F}) \subset \mathcal{M}$. Conversely if $f^{-1}(\mathcal{E}) \subset \mathcal{M}$ then $\sigma\left(f^{-1}(\mathcal{E})\right) \subset \mathcal{M}$ and so making use of Lemma 6.3,

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

■

Corollary 6.11. *Suppose that $(X, \mathcal{M})$ is a measurable space. Then the following conditions on a function $f : X \to \mathbb{R}$ are equivalent:*

1. *f is $(\mathcal{M}, \mathcal{B}_\mathbb{R})$ – measurable,*
2. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$,*
3. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{Q}$,*
4. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Exercise 6.2. Prove Corollary 6.11. **Hint:** See Exercise 3.7.

Exercise 6.3. If $\mathcal{M}$ is the σ – algebra generated by $\mathcal{E} \subset 2^X$, then $\mathcal{M}$ is the union of the σ – algebras generated by countable subsets $\mathcal{F} \subset \mathcal{E}$.

Exercise 6.4. Let $(X, \mathcal{M})$ be a measure space and $f_n : X \to \mathbb{R}$ be a sequence of measurable functions on X. Show that $\{x : \lim_{n\to\infty} f_n(x) \text{ exists } \textbf{in } \mathbb{R}\} \in \mathcal{M}$.

Exercise 6.5. Show that every monotone function $f : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}_\mathbb{R}, \mathcal{B}_\mathbb{R})$ – measurable.

Definition 6.12. *Given measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ and a subset $A \subset X$. We say a function $f : A \to Y$ is measurable iff f is $\mathcal{M}_A/\mathcal{F}$ – measurable.*

Proposition 6.13 (Localizing Measurability). *Let $(X, \mathcal{M})$ and $(Y, \mathcal{F})$ be measurable spaces and $f : X \to Y$ be a function.*

1. *If f is measurable and $A \subset X$ then $f|_A : A \to Y$ is 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 6.14. *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 6.15. *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 $-\infty < a < b < \infty$.

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

$$G^{-1}((0, x_0]) = \{y : G(y) \leq x_0\} = (0, F(x_0)] \cap \mathbb{R}, \tag{6.3}$$

see Figure 6.2.

To give a formal proof of Eq. (6.3), $G(y) = \inf\{x : F(x) \geq y\} \leq x_0$, 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

$$\{G \leq x_0\} \subset (0, F(x_0)] \cap (0, 1).$$

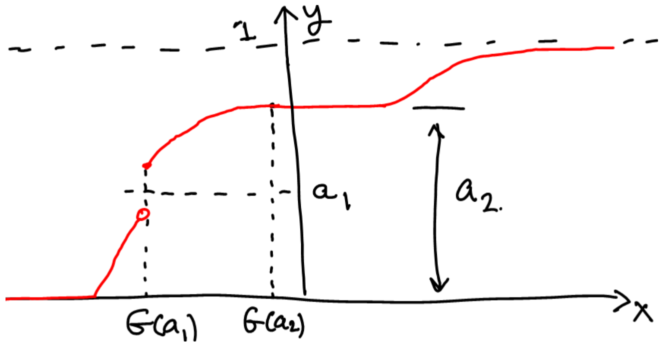


Fig. 6.1. A pictorial definition of G.

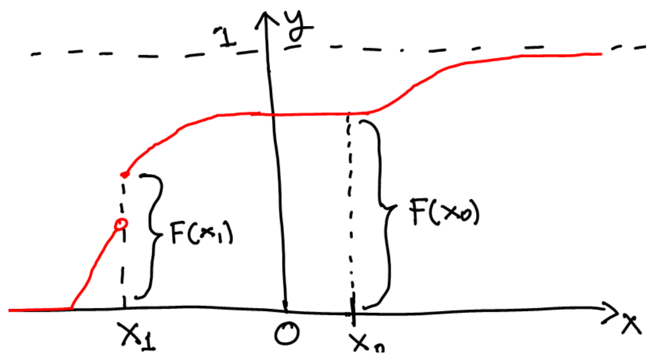


Fig. 6.2. As can be seen from this picture, $G(y) \le x_0$ iff $y \le F(x_0)$ and similarly, $G(y) \le x_1$ iff $y \le x_1$.

For the converse, if $y \le F(x_0)$ then $G(y) = \inf\{x : F(x) \ge y\} \le x_0$, i.e. $y \in \{G \le x_0\}$. Indeed, $y \in G^{-1}((-\infty, x_0])$ iff $G(y) \le x_0$. Observe that

$$G(F(x_0)) = \inf\{x : F(x) \ge F(x_0)\} \le x_0$$

and hence $G(y) \le x_0$ whenever $y \le F(x_0)$. 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,

$$\begin{aligned}(G_* m)((-\infty, x]) &= m\left(G^{-1}((-\infty, x])\right) = m(\{y \in (0,1) : G(y) \le x\}) \\ &= m((0, F(x)] \cap (0,1)) = F(x).\end{aligned}$$

See section 2.5.2 on p. 61 of Resnick for more details. ∎

Theorem 6.16 (Durret's Version). *Given a distribution function, $F : \mathbb{R} \to [0,1]$ let $Y : (0,1) \to \mathbb{R}$ be defined (see Figure 6.3) by,*

$$Y(x) := \sup\{y : F(y) < x\}.$$

Then $Y : (0,1) \to \mathbb{R}$ is Borel measurable and $Y_* m = \mu_F$ where μ_F is the unique measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $-\infty < a < b < \infty$.

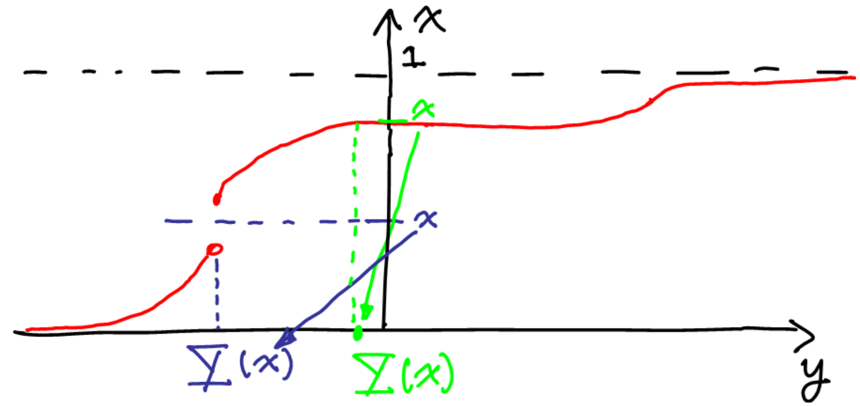


Fig. 6.3. A pictorial definition of $Y(x)$.

Proof. Since $Y : (0,1) \to \mathbb{R}$ is a non-decreasing function, Y is measurable. Also observe, if $y < Y(x)$, then $F(y) < x$ and hence,

$$F(Y(x)-) = \lim_{y \uparrow Y(x)} F(y) \le x.$$

For $y > Y(x)$, we have $F(y) \ge x$ and therefore,

$$F(Y(x)) = F(Y(x)+) = \lim_{y \downarrow Y(x)} F(y) \ge x$$

and so we have shown

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

We will now show

$$\{x \in (0,1) : Y(x) \le y_0\} = (0, F(y_0)] \cap (0,1). \tag{6.4}$$

For the inclusion "$\subset$," if $x \in (0,1)$ and $Y(x) \le y_0$, then $x \le F(Y(x)) \le F(y_0)$, i.e. $x \in (0, F(y_0)] \cap (0,1)$. Conversely if $x \in (0,1)$ and $x \le F(y_0)$ then (by definition of $Y(x)$) $y_0 \ge Y(x)$.

From the identity in Eq. (6.4), it follows that Y is measurable and

$$(Y_* m)((-\infty, y_0)) = m\left(Y^{-1}(-\infty, y_0)\right) = m((0, F(y_0)] \cap (0,1)) = F(y_0).$$

Therefore, $Law(Y) = \mu_F$ as desired. ∎

Lemma 6.17 (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 6.18 (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 6.19. Suppose that Y is a finite set, $\mathcal{F} = 2^Y$, and $X = Y^N$ for some $N \in \mathbb{N}$. Let $\pi_i : Y^N \to Y$ be the projection maps, $\pi_i(y_1, \dots, y_N) = y_i$. Then, as the reader should check,

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

Proposition 6.20. *Assuming the notation in Definition 6.18 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 6.17. ($\Leftarrow$) 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 6.21. *A function $f : X \to Y$ between two topological spaces is **Borel measurable** if $f^{-1}(\mathcal{B}_Y) \subset \mathcal{B}_X$.*

Proposition 6.22. *Let X and Y be two topological spaces and $f : X \to Y$ be a continuous function. Then f is Borel measurable.*

Proof. Using Lemma 6.3 and $\mathcal{B}_Y = \sigma(\tau_Y)$,

$$f^{-1}(\mathcal{B}_Y) = f^{-1}(\sigma(\tau_Y)) = \sigma(f^{-1}(\tau_Y)) \subset \sigma(\tau_X) = \mathcal{B}_X.$$

∎

Example 6.23. For $i = 1, 2, \dots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be defined by $\pi_i(x) = x_i$. Then each π_i is continuous and therefore $\mathcal{B}_{\mathbb{R}^n}/\mathcal{B}_{\mathbb{R}}$ – measurable.

Lemma 6.24. *Let $\mathcal{E}$ denote the collection of open rectangle in $\mathbb{R}^n$, then $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{E})$. We also have that $\mathcal{B}_{\mathbb{R}^n} = \sigma(\pi_1, \dots, \pi_n)$ and in particular, $A_1 \times \dots \times A_n \in \mathcal{B}_{\mathbb{R}^n}$ whenever $A_i \in \mathcal{B}_{\mathbb{R}}$ for $i = 1, 2, \dots, n$. Therefore $\mathcal{B}_{\mathbb{R}^n}$ may be described as the σ algebra generated by $\{A_1 \times \dots \times A_n : A_i \in \mathcal{B}_{\mathbb{R}}\}$.*

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 $i = 1, 2, \dots, n$ and let

$$(a, b) = (a_1, b_1) \times \dots \times (a_n, b_n). \tag{6.5}$$

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. (6.5), 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 6.25. *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 6.24 and Proposition 6.20. ■

Corollary 6.26. *Let $(X, \mathcal{M})$ be a measurable space and $f, g : X \to \mathbb{C}$ be $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable functions. Then $f \pm g$ and $f \cdot g$ are also $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $A_\pm : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ and $M : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$ by $F(x) = (f(x), g(x))$, $A_\pm(w, z) = w \pm z$ and $M(w, z) = wz$. Then $A_\pm$ and M are continuous and hence $(\mathcal{B}_{\mathbb{C}^2}, \mathcal{B}_{\mathbb{C}})$ – measurable. Also F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}^2})$ – measurable since $\pi_1 \circ F = f$ and $\pi_2 \circ F = g$ are $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ – measurable. Therefore $A_\pm \circ F = f \pm g$ and $M \circ F = f \cdot g$, being the composition of measurable functions, are also measurable. ■

Lemma 6.27. *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*

$$F(x) := \begin{cases} \frac{1}{f(x)} & \text{if } f(x) \neq 0 \\ \alpha & \text{if } f(x) = 0 \end{cases}$$

is measurable.

Proof. Define $i : \mathbb{C} \to \mathbb{C}$ by

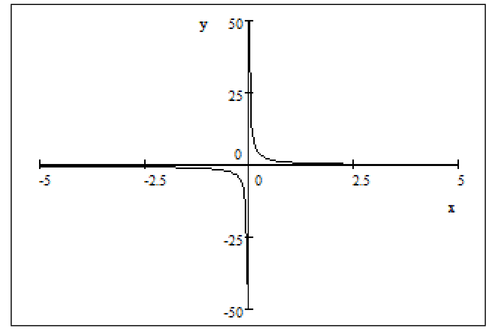
$$i(z) = \begin{cases} \frac{1}{z} & \text{if } z \neq 0 \\ 0 & \text{if } z = 0. \end{cases}$$

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 6.28. For the real case of Lemma 6.27, define i as above but now take z to real. From the plot of i, Figure 6.28, 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

$$\mathcal{B}_{\bar{\mathbb{R}}} := \sigma(\{[a, \infty] : a \in \mathbb{R}\}). \tag{6.6}$$

Proposition 6.29 (The Structure of $\mathcal{B}_{\bar{\mathbb{R}}}$). *Let $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\bar{\mathbb{R}}}$ be as above, then*

$$\mathcal{B}_{\bar{\mathbb{R}}} = \{A \subset \bar{\mathbb{R}} : A \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}. \tag{6.7}$$

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. (6.7). ∎

The proofs of the next two corollaries are left to the reader, see Exercises 6.6 and 6.7.

Corollary 6.30. *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 6.31. *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 6.6. Prove Corollary 6.30 noting that the equivalence of items 1. – 3. is a direct analogue of Corollary 6.11. Use Proposition 6.29 to handle item 4.

Exercise 6.7. Prove Corollary 6.31.

Proposition 6.32 (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 6.33. *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 6.34. *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 6.32 implies $f_\pm$ are measurable. Conversely if $f_\pm$ are measurable then so is $f = f_+ - f_-$. ∎

Definition 6.35. *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

$$\varphi = \sum_{i=1}^n \lambda_i 1_{A_i} \text{ with } A_i \in \mathcal{M} \text{ and } \lambda_i \in \mathbb{F}. \tag{6.8}$$

Indeed, take $\lambda_1, \lambda_2, \dots, \lambda_n$ to be an enumeration of the range of φ and $A_i = \varphi^{-1}(\{\lambda_i\})$. Note that this argument shows that any simple function may be written intrinsically as

$$\varphi = \sum_{y \in \mathbb{F}} y 1_{\varphi^{-1}(\{y\})}. \tag{6.9}$$

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

Theorem 6.36 (Approximation Theorem). *Let $f : X \to [0, \infty]$ be measurable and define, see Figure 6.4,*

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

then $\varphi_n \le f$ for all n, $\varphi_n(x) \uparrow f(x)$ for all $x \in X$ and $\varphi_n \uparrow f$ uniformly on the sets $X_M := \{x \in X : f(x) \le M\}$ with $M < \infty$.

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

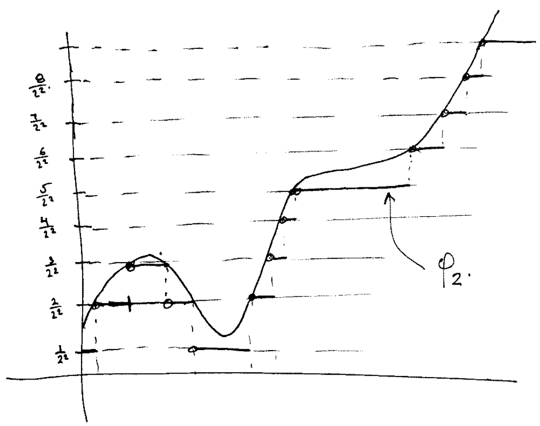


Fig. 6.4. Constructing simple functions approximating a function, $f : X \to [0, \infty]$.

Proof. Since

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

if $x \in f^{-1}\left((\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}]\right)$ then $\varphi_n(x) = \varphi_{n+1}(x) = \frac{2k}{2^{n+1}}$ and if $x \in f^{-1}\left((\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}]\right)$ then $\varphi_n(x) = \frac{2k}{2^{n+1}} < \frac{2k+1}{2^{n+1}} = \varphi_{n+1}(x)$. Similarly

$$(2^n, \infty] = (2^n, 2^{n+1}] \cup (2^{n+1}, \infty],$$

and so for $x \in f^{-1}((2^{n+1}, \infty])$, $\varphi_n(x) = 2^n < 2^{n+1} = \varphi_{n+1}(x)$ and for $x \in f^{-1}((2^n, 2^{n+1}])$, $\varphi_{n+1}(x) \ge 2^n = \varphi_n(x)$. Therefore $\varphi_n \le \varphi_{n+1}$ for all n. It is clear by construction that $\varphi_n(x) \le f(x)$ for all x and that $0 \le f(x) - \varphi_n(x) \le 2^{-n}$ if $x \in X_{2^n}$. 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

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

and clearly $|\varphi_n| = \varphi_n^+ + \varphi_n^- \uparrow f_+ + f_- = |f|$ and $\varphi_n = \varphi_n^+ - \varphi_n^- \to f_+ - f_- = f$ as $n \to \infty$. Now suppose that $f : X \to \mathbb{C}$ is measurable. We may now choose simple function u_n and v_n such that $|u_n| \uparrow |\mathrm{Re}\, f|$, $|v_n| \uparrow |\mathrm{Im}\, f|$, $u_n \to \mathrm{Re}\, f$ and $v_n \to \mathrm{Im}\, f$ as $n \to \infty$. Let $\varphi_n = u_n + iv_n$, then

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

and $\varphi_n = u_n + iv_n \to \mathrm{Re}\, f + i\, \mathrm{Im}\, f = f$ as $n \to \infty$. ■

6.2 Factoring Random Variables

Lemma 6.37. *Suppose that $(\mathbb{Y}, \mathcal{F})$ is a measurable space and $Y : \Omega \to \mathbb{Y}$ is a map. Then to every $(\sigma(Y), \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, $H : \Omega \to \bar{\mathbb{R}}$, there is a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function $h : \mathbb{Y} \to \bar{\mathbb{R}}$ such that $H = h \circ Y$.*

Proof. First suppose that $H = 1_A$ where $A \in \sigma(Y) = Y^{-1}(\mathcal{F})$. Let $B \in \mathcal{F}$ such that $A = Y^{-1}(B)$ then $1_A = 1_{Y^{-1}(B)} = 1_B \circ Y$ and hence the lemma is valid in this case with $h = 1_B$. More generally if $H = \sum a_i 1_{A_i}$ is a simple function, then there exists $B_i \in \mathcal{F}$ such that $1_{A_i} = 1_{B_i} \circ Y$ and hence $H = h \circ Y$ with $h := \sum a_i 1_{B_i}$ – a simple function on $\bar{\mathbb{R}}$.

For a general $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function, H, from $\Omega \to \bar{\mathbb{R}}$, choose simple functions H_n converging to H. Let $h_n : \mathbb{Y} \to \bar{\mathbb{R}}$ be simple functions such that $H_n = h_n \circ Y$. Then it follows that

$$H = \lim_{n\to\infty} H_n = \limsup_{n\to\infty} H_n = \limsup_{n\to\infty} h_n \circ Y = h \circ Y$$

where $h := \limsup_{n\to\infty} h_n$ – a measurable function from $\mathbb{Y}$ to $\bar{\mathbb{R}}$. ■

The following is an immediate corollary of Proposition 6.20 and Lemma 6.37.

Corollary 6.38. *Let X and A be sets, and suppose for $\alpha \in A$ we are give a measurable space $(Y_\alpha, \mathcal{F}_\alpha)$ and a function $f_\alpha : X \to Y_\alpha$. Let $Y := \prod_{\alpha \in A} Y_\alpha$, $\mathcal{F} := \otimes_{\alpha \in A} \mathcal{F}_\alpha$ be the product σ – algebra on Y and $\mathcal{M} := \sigma(f_\alpha : \alpha \in A)$ be the smallest σ – algebra on X such that each f_α is measurable. Then the function $F : X \to Y$ defined by $[F(x)]_\alpha := f_\alpha(x)$ for each $\alpha \in A$ is $(\mathcal{M}, \mathcal{F})$ – measurable and a function $H : X \to \bar{\mathbb{R}}$ is $(\mathcal{M}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable iff there exists a $(\mathcal{F}, \mathcal{B}_{\bar{\mathbb{R}}})$ – measurable function h from Y to $\bar{\mathbb{R}}$ such that $H = h \circ F$.*

References

1. Richard F. Bass, *Probabilistic techniques in analysis*, Probability and its Applications (New York), Springer-Verlag, New York, 1995. MR MR1329542 (96e:60001)
2. ______, *The Doob-Meyer decomposition revisited*, Canad. Math. Bull. **39** (1996), no. 2, 138–150. MR MR1390349 (97b:60075)
3. ______, *Diffusions and elliptic operators*, Probability and its Applications (New York), Springer-Verlag, New York, 1998. MR MR1483890 (99h:60136)
4. Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995, A Wiley-Interscience Publication. MR MR1324786 (95k:60001)
5. R. M. Blumenthal and R. K. Getoor, *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York, 1968. MR MR0264757 (41 #9348)
6. K. L. Chung and R. J. Williams, *Introduction to stochastic integration*, second ed., Probability and its Applications, Birkhäuser Boston Inc., Boston, MA, 1990. MR MR1102676 (92d:60057)
7. Claude Dellacherie, *Capacités et processus stochastiques*, Springer-Verlag, Berlin, 1972, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67. MR MR0448504 (56 #6810)
8. J. Dieudonné, *Foundations of modern analysis*, Academic Press, New York, 1969, Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I. MR MR0349288 (50 #1782)
9. J. L. Doob, *Stochastic processes*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR MR1038526 (91d:60002)
10. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996. MR MR1609153 (98m:60001)
11. ______, *Stochastic calculus*, Probability and Stochastics Series, CRC Press, Boca Raton, FL, 1996, A practical introduction. MR MR1398879 (97k:60148)
12. Evgenii B. Dynkin and Aleksandr A. Yushkevich, *Markov processes: Theorems and problems*, Translated from the Russian by James S. Wood, Plenum Press, New York, 1969. MR MR0242252 (39 #3585a)
13. Masatoshi Fukushima, Yōichi Ōshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR MR1303354 (96f:60126)
14. Robert D. Gordon, *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument*, Ann. Math. Statistics **12** (1941), 364–366. MR MR0005558 (3,171e)
15. Nobuyuki Ikeda and Shinzo Watanabe, *Stochastic differential equations and diffusion processes*, second ed., North-Holland Mathematical Library, vol. 24, North-Holland Publishing Co., Amsterdam, 1989. MR MR1011252 (90m:60069)
16. Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 1987. MR MR959133 (89k:60044)
17. ______, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR MR1943877 (2003j:60001)
18. Svante Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997. MR MR1474726 (99f:60082)
19. Shizuo Kakutani, *On equivalence of infinite product measures*, Ann. of Math. (2) **49** (1948), 214–224. MR MR0023331 (9,340e)

20. Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR MR1876169 (2002m:60002)
21. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR MR1121940 (92h:60127)
22. Nicolai Krylov, *A simple proof of a result of A. Novikov*, 2002.
23. Hiroshi Kunita, *Stochastic flows and stochastic partial differential equations*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986) (Providence, RI), Amer. Math. Soc., 1987, pp. 1021–1031. MR MR934304 (89g:60194)
24. ______, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990. MR MR1070361 (91m:60107)
25. Hui Hsiung Kuo, *Gaussian measures in Banach spaces*, Springer-Verlag, Berlin, 1975, Lecture Notes in Mathematics, Vol. 463. MR 57 #1628
26. Shigeo Kusuoka and Daniel W. Stroock, *Precise asymptotics of certain Wiener functionals*, J. Funct. Anal. **99** (1991), no. 1, 1–74. MR MR1120913 (93a:60085)
27. Gregory F. Lawler, *Introduction to stochastic processes*, Chapman & Hall Probability Series, Chapman & Hall, New York, 1995. MR MR1372946 (97a:60001)
28. Terry Lyons and Zhongmin Qian, *System control and rough paths*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2002, Oxford Science Publications. MR MR2036784 (2005f:93001)
29. Zhi Ming Ma and Michael Röckner, *Introduction to the theory of (nonsymmetric) Dirichlet forms*, Universitext, Springer-Verlag, Berlin, 1992. MR MR1214375 (94d:60119)
30. Henry P. McKean, *Stochastic integrals*, AMS Chelsea Publishing, Providence, RI, 2005, Reprint of the 1969 edition, with errata. MR MR2169626 (2006d:60003)
31. Michel Métivier, *Semimartingales*, de Gruyter Studies in Mathematics, vol. 2, Walter de Gruyter & Co., Berlin, 1982, A course on stochastic processes. MR MR688144 (84i:60002)
32. J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR MR1600720 (99c:60144)
33. James Norris, *Simplified Malliavin calculus*, Séminaire de Probabilités, XX, 1984/85, Lecture Notes in Math., vol. 1204, Springer, Berlin, 1986, pp. 101–130.
34. Yuval Peres, *An invitation to sample paths of brownian motion*, stat-www.berkeley.edu/ peres/bmall.pdf (2001), 1–68.
35. ______, *Brownian motion*, http://www.stat.berkeley.edu/users/peres/bmbook.pdf (2006), 1–272.
36. Philip Protter, *Stochastic integration and differential equations*, Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 1990, A new approach. MR MR1037262 (91i:60148)
37. Philip E. Protter, *Stochastic integration and differential equations*, second ed., Applications of Mathematics (New York), vol. 21, Springer-Verlag, Berlin, 2004, Stochastic Modelling and Applied Probability. MR MR2020294 (2005k:60008)
38. ______, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR MR2273672
39. Michael Reed and Barry Simon, *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975. MR 58 #12429b
40. ______, *Methods of modern mathematical physics. I*, second ed., Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Functional analysis. MR 85e:46002
41. Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999. MR MR1725357 (2000h:60050)
42. L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR 2001g:60188
43. ______, *Diffusions, Markov processes, and martingales. Vol. 2*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Itô calculus, Reprint of the second (1994) edition. MR MR1780932 (2001g:60189)
44. Sheldon M. Ross, *Stochastic processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1983, Lectures in Mathematics, 14. MR MR683455 (84m:60001)
45. H. L. Royden, *Real analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR MR1013117 (90g:00004)
46. Robert Schatten, *Norm ideals of completely continuous operators*, Second printing. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 27, Springer-Verlag, Berlin, 1970. MR 41 #2449
47. Michael Sharpe, *General theory of Markov processes*, Pure and Applied Mathematics, vol. 133, Academic Press Inc., Boston, MA, 1988. MR MR958914 (89m:60169)
48. Barry Simon, *Trace ideals and their applications*, London Mathematical Society Lecture Note Series, vol. 35, Cambridge University Press, Cambridge, 1979. MR 80k:47048
49. ______, *Functional integration and quantum physics*, second ed., AMS Chelsea Publishing, Providence, RI, 2005. MR MR2105995 (2005f:81003)
50. S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes in Mathematics, vol. 7, New York University Courant Institute of Mathematical Sciences, New York, 2001. MR MR1852999 (2003a:60001)
51. John B. Walsh, *An introduction to stochastic partial differential equations*, École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., vol. 1180, Springer, Berlin, 1986, pp. 265–439. MR MR876085 (88a:60114)