MATH 6605
Probability Theory

Joe Tran

Winter 2025

Preface

These are the first edition of these lecture notes for MATH 6605 (Probability Theory). Consequently, there may be several typographical errors, missing exposition on necessary background, and more advanced topics for which there will not be time in class to cover. Future iterations of these notes will hopefully be fairly self-contained provided one has the necessary background. If you come across any typos, errors, omissions, or unclear expositions, please feel free to contact me so that I may continually improve these notes.

Contents

Chapter 1

The Need For Measure Theory

This introductory section is directed primarily to those who have some familiarity with undergraduate level probability theory, and who may be unclear as to why it is necessary to introduce measure theory and other mathematical difficulties in order to study probability theory in a rigorous manner.

We attempt to illustrate the limitations of undergraduate-level probability in two different ways: the restrictions on the kinds of random variables it allows, and the question of what sets can probabilities defined on them.

1.1 Various Kinds of Random Variables

The reader with undergraduate level probability will be comfortable with the statement like, "Let X be a random variable which has the Poisson(5) distribution." The reader will know that this means that X takes as its value a "random" nonnegative integer such that the integer $k \geq 0$ chosen with probability

$$P(X = k) = \frac{e^{-5}5^k}{k!}$$

The expected value of, say, X^2, can then be computed as

$$\mathcal{E}(X^2) = \sum_{k=0}^{\infty} x^2 \frac{e^{-5}5^k}{k!}$$

X is an example of a *discrete random variable*.

Similarly, the reader will be familiar with the statement like, "Let Y be a random variable which has the Normal$(0, 1)$ distribution." This means that the probability that Y lies between two real numbers $a < b$ is given by the integral

$$P(a \leq Y \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

The expected value, of say, Y^2, can then by computed as

$$\mathcal{E}(Y^2) = \int_{-\infty}^{+\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Y is an example of an *absolutely continuous random variable*, or simply, *continuous random variable*.

But now suppose we introduce a new random variable, Z as follows: We let X and Y be as above, and then flip an (independent) fair coin. If the coin comes up heads, then we set $Z = X$, while if it comes up tails, we set $Z = Y$. In symbols,

$$P(Z = X) = P(Z = Y) = \frac{1}{2}$$

Then what sort of random variable is Z? It is not discrete since it can take on an uncountable number of different values, and it is not absolutely continuous for certain values of z, in particular, when z is a nonnegative integer, we have $P(Z = z) > 0$. So how can we study the random variable Z, and how can we compute the expect value of Z^2.

The correct response to this question, is that the division of random variables into discrete versus absolutely continuous is artificial. Instead, measure theory allows us to give a common definition of expected value, which applies equally well to discrete random variables to a continuous random variables to combinations of them, and other kinds of random variables not yet imagined.

1.2 The Uniform Distribution and Non-Measurable Sets

In undergraduate-level probability, continuous random variables are often studied in detail. However, a closer examination suggests that perhaps such random variables are not completely understood at all.

To take the simplest case, suppose that X is a random variable which has the uniform distribution on the unit interval $[0, 1]$. In symbols, $X \sim \text{Uniform}[0, 1]$. What does this precisely mean.

This means that

$$\begin{aligned} P(0 \leq X \leq 1) &= 1 \\ P(0 \leq X \leq 1/2) &= \frac{1}{2} \\ P(3/4 \leq X \leq 7/8) &= \frac{1}{8} \end{aligned}$$

and so on. In general, this means

$$P(a \leq X \leq b) = b - a$$

whenever $0 \leq a \leq b \leq 1$ with the same formula holding if $\leq$ is replaced with $<$. Indeed,

$$P([a, b]) = P((a, b]) = P([a, b)) = P((a, b)) = b - a \tag{1.1}$$

for all $0 \leq a \leq b \leq 1$. In other words, the probability that X lies in any interval contained in $[0, 1]$ is simply the length of the interval. Similarly, this means

$$P([1/4, 1/2] \cup [2/3, 5/6]) = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

and in general, if A and B are disjoint subsets of $[0,1]$, then

$$P(A \cup B) = P(A) + P(B) \tag{1.2}$$

Equation (1.2) is called *finite additivity.*

Similarly, to allow for countable operations (such as limits, which are extremely important in probability theory), we could extend (1.2) to the case of a countably infinite number of disjoint subsets: if $(A_n)_{n=1}^{\infty}$ is a collection of disjoint subsets of $[0,1]$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \tag{1.3}$$

Equation (1.3) is called *countable additivity.*

Note that we do not extend (1.3) to *uncountable additivity.* Indeed, if we did, then we would expect that

$$P([0,1]) = \sum_{x \in [0,1]} P(\{x\})$$

which is absurd, since the left side is equal to 1, but the right side is equal to 0, but note there is no contradiction to (1.3) since $[0,1]$ is not even countable.

Similarly, to reflect the fact that X is "uniform" on the interval $[0,1]$, the probability that X lies in some subset should be unaffected by "shifting" (with wrap-around) the subset by a fixed amount. That is, if for each subset $A \subseteq [0,1]$, we define the *r-shift of A* by

$$A \oplus r = \{a + r : a \in A, a + r \leq 1\} \cup \{a + r - 1 : a \in A, a + r > 1\} \tag{1.4}$$

then we should have

$$P(A \oplus r) = P(A) \tag{1.5}$$

for all $0 \leq r \leq 1$.

Now suppose we ask, what is the probability that X is rational. What is the probability that X^n is rational for some positive integer n? What is the probability that X is *algebraic*, i.e. the solution to some polynomial with integer coefficients? Can we compute these things? More fundamentally, are all probabilities such as these necessarily even *defined*? That is, does $P(A)$ (i.e. the probability that X lies in the subset A) even make sense for every possible subset $A \subseteq [0,1]$?

The answer to the last question is no, as the following proposition shows.

Proposition 1.2.1

There does not exist a definition of $P(A)$, defined for all subsets $A \subseteq [0,1]$ satisfying (1.1), (1.3) and (1.5).

Proof.

Suppose for a contradiction that $P(A)$ could be defined for any subset $A \subseteq [0,1]$. Define an equivalence relation "$\sim$" on $[0,1]$ as follows: $x \sim y$ if and only if $y - x \in \mathbb{Q}$. This relation partitions the interval $[0,1]$ into a disjoint union of equivalence classes. This relation partitions the interval $[0,1]$ into a disjoint union of equivalence classes. Let H be a subset of $[0,1]$ consisting of precisely one element from each equivalence class (such H exists by the Axiom of Choice). For definiteness, assume that $0 \notin H$ (say, if $0 \in H$, replace by $1/2$).

Now since H contains an element of each equivalence class, we see that each point in $(0,1]$ is contained in the union $\bigcup_{r \in [0,1) \cap \mathbb{Q}} (H \oplus r)$ of shifts of H. Furthermore, since H contains just *one* point from each equivalence class, we see that the sets $H \oplus r$ for $r \in [0,1)$ rational, are all disjoint.

But then by (1.3), we have

$$P((0,1]) = \sum_{r \in [0,1) \cap \mathbb{Q}} P(H \oplus r)$$

Shift-invariance (1.5) implies that $P(H \oplus r) = P(H)$ and so

$$1 = P((0,1]) = \sum_{r \in [0,1) \cap \mathbb{Q}} P(H)$$

which is absurd, since a countably infinite sum of the same quantity repeated can only equal to 0, ∞, or $-\infty$, but it can never equal to 1. ∎

This proposition says that if we want our probabilities to satisfy reasonable properties, then we *cannot* define them for all possible subsets of $[0,1]$. Rather, we must *restrict* their definition to certain "measurable" sets. This is the motivation for the next chapter.

Remark 1.2.2

The existence of problematic sets like H above turns out to be equivalent to the axiom of choice. In particular, we can never define such sets explicitly—only implicitly by the axiom of choice as in the proof in Proposition 1.2.1.

Chapter 2

Probability Triples

In this section, we consider the probability triples and how to construct them. In light of the previous chapter, we see that to study probability theory properly, it will be necessary to keep track of which subsets A have a probability $P(A)$ defined for them.

2.1 Basic Definitions

Definition 2.1.1

We define a *probability triple* or (*probability*) *measure space* or *probability space* to be a triple $(\Omega, \mathcal{A}, P)$, where Ω is the *sample space* that is any nonempty set, $\mathbb{F}$ is a σ*-algebra* which is a collection of subsets of Ω, containing Ω itself and the empty set $\emptyset$, and closed under the formation of complements and countable unions and countable intersections, and the *probability measure* P is a mapping from $\mathcal{A}$ to $[0,1]$ with $P(\emptyset) = 0$ and $P(\Omega) = 1$, such a P is countably additive as in (1.3)

This definition will be in constant use throughout the text. Furthermore, it contains a number of subtle points. Thus, we pause to make a few observations.

The σ-algebra $\mathcal{A}$ is the collection of all *events* or *measurable sets*. These are the subsets $A \subseteq \Omega$ for which $P(A)$ is well-defined. From Proposition 1.2.1 that in general $\mathcal{A}$ may not contain all subsets of Ω, though we still expect it to contain most of the subsets that come up naturally.

To say that $\mathcal{A}$ is closed under the formation of complements and countable unions and countable intersections means

(i) For all $A \in \mathcal{A}$, if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

(ii) For any countable collection $(A_n)_{n=1}^{\infty}$ of Ω, if for all $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

(iii) For any countable collection $(A_n)_{n=1}^{\infty}$ of ω, if for all $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Like for countable additivity, the reason why we require $\mathcal{A}$ to be closed under countable operations is to allow for taking limits, etc., when studying probability theory. Also, like for additivity, we cannot extend the definition to require that $\mathcal{A}$ is closed under uncountable unions; in this case, for the example in Section 1.2, $\mathcal{A}$ would contain every subset A, since every subset can be written as $A = \bigcup_{x\in A}\{x\}$ and since the singleton sets $\{x\}$ are all in $\mathcal{A}$.

There are some redundancy in the definition above. For example, it follows from De Morgan's Law that if $\mathcal{A}$ is closed under complement and countable unions, then it is automatically closed under countable intersections. Similarly, it follows from countable additivity that we must have $P(\emptyset) = 0$ and that (once we know that $P(\Omega) = 1$ and $P(A) \geq 0$ for all $A \in \mathcal{A}$), we must have $P(A) \leq 1$.

More generally, from additivity, we have $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$, whence

$$P(A^c) = 1 - P(A) \tag{2.1}$$

a fact that will be used often. Similarly, if $A \subseteq B$, then since $B = A \sqcup (B \setminus A)$, where $\sqcup$ means disjoint union, we have that

$$P(B) = P(A) + P(B \setminus A) \geq P(A)$$

and thus,

$$P(A) \leq P(B) \tag{2.2}$$

whenever $A \subseteq B$, which is the *monotonicity property of probability measures.*

Also, if $A, B \in \mathcal{A}$, then

$$\begin{aligned}
P(A \cup B) &= P((A \setminus B) \sqcup (B \setminus A) \sqcup (A \cap B)) \\
&= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\
&= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}$$

the *principle of inclusion-exclusion.*

Finally, if $A_0 = \emptyset$ and for any sequence $(A_n)_{n=1}^{\infty}$ of $\mathcal{A}$ (disjoint or not), we have by the countable additivity and monotonicity that

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigsqcup_{n=1}^{\infty} (A_n \setminus A_{n-1})\right) = \sum_{n=1}^{\infty} P(A_n \setminus A_{n-1}) \leq \sum_{n=1}^{\infty} P(A_n)$$

which is the *countable subadditivity* property of probability measures.

2.2 Constructing Probability Triples

We clarify the definition of Section 2.1 with a simple example.

Example 2.2.1

Let us again consider Poisson(5) distribution that was mentioned in Section 1.1. In this case, the sample Ω would consist of all nonnegative integers: $\Omega = \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Also, the

σ-algebra $\mathcal{A}$ would consist of all subsets of $\mathbb{N}_0$. Finally, the probability measure P would be defined as, for any $A \in \mathcal{A}$,

$$P(A) = \sum_{a \in A} \frac{e^{-5} 5^a}{a!}$$

It is elementary to check that $\mathcal{A}$ is a σ-algebra, as it contains all subsets of $\mathbb{N}_0$, so it is closed under any set operations, and it is also elementary to check that P is a probability measure defined on $\mathcal{A}$ (the additivity following since if A and B are disjoint, then $\sum_{x \in A \cup B}$ is the same as $\sum_{x \in A} + \sum_{x \in B}$).

So in the case of Poisson(5), we see that it is entirely straightforward to construct an appropriate triple $(\mathbb{N}_0, \mathcal{A}, P)$. The construction is similarly straightforward for any discrete probability space, i.e. any space for which Ω is finite or countable. We record this as follows.

Theorem 2.2.2

Let Ω be a finite or countable nonempty set. Let $p : \Omega \to [0, 1]$ be any function satisfying $\sum_{\omega \in \Omega} p(\omega) = 1$. Then there exists a valid probability triple $(\Omega, \mathcal{A}, P)$ where $\mathcal{A}$ is the collection of all subsets of Ω, and for $A \in \mathcal{A}$, $P(A) = \sum_{a \in A} p(a)$.

Example 2.2.3

Let Ω be any finite nonempty set, $\mathcal{A}$ be the collection of all subsets of Ω, and P be the probability measure defined as

$$P(A) = \frac{|A|}{|\Omega|}$$

for all $A \in \mathcal{A}$, where $|A|$ denotes the cardinality of the set A. Then $(\Omega, \mathcal{A}, P)$ is a valid probability triple, called the *uniform distribution on* Ω, denoted by Uniform(Ω).

However, if a sample space is not countable, then the situation is considerably more complex, as seen in Section 1.2. How can we formally define a probability triple $(\Omega, \mathcal{A}, P)$ which corresponds to say, the Uniform($[0, 1]$) distribution?

It is clear that we should choose $\Omega = [0, 1]$, but what about $\mathcal{A}$? We know from Proposition 1.2.1 that $\mathcal{A}$ cannot contain all intervals of Ω, but it should contain all the intervals $[a, b]$, $[a, b)$, etc. That is, we must have $\mathcal{A} \supset \mathcal{B}$, where

$$\mathcal{B} = \{\text{all intervals contained in } [0, 1]\}$$

and where "intervals" is understood to include the open, closed, half-open, singleton, and empty intervals.

Exercise 2.2.4

Prove that the above collection $\mathcal{B}$ is a *semi-algebra* of subsets of Ω. That is, show

(i) $\Omega, \emptyset \in \mathcal{B}$

(ii) $\mathcal{B}$ is closed under finite intersection

(iii) The complement of any element of $\mathcal{B}$ is equal to a finite disjoint union of elements of $\mathcal{B}$.

Since $\mathcal{B}$ is only a semi-algebra, how can we create a σ-algebra? As a first try, we might consider

$$\mathcal{C} = \{\text{all finite unions of elements of } \mathcal{B}\} \tag{2.3}$$

After all, by additivity, we already know how to define P on $\mathcal{C}$. However, $\mathcal{C}$ is not a σ-algebra.

Exercise 2.2.5

(a) Prove that $\mathcal{C}$ is an *algebra* (or, *field*) of subsets of Ω, that is,

  (i) $\mathcal{C}$ contains Ω and $\emptyset$

  (ii) $\mathcal{C}$ is closed under the formation of complements and of finite unions and intersections.

(b) Prove that $\mathcal{C}$ is not a σ-algebra.

As a second try, we might even consider

$$\mathcal{D} = \{\text{all finite or countable unions of elements of } \mathcal{B}\} \tag{2.4}$$

Unfortunately, $\mathcal{D}$ is not a σ-algebra.

Thus, the construction of $\mathcal{A}$ and of P presents serious challenges. To deal with them, we next prove a very general theorem about constructing probability triples.

2.3 The Extension Theorem*

The following theorem is of fundamental importance in constructing complicated probability triples. Recall the definition of *semi-algebra* from Exercise 2.2.4.

Theorem 2.3.1: The Extension Theorem

Let $\mathcal{B}$ be a semi-algebra of subsets of Ω. Let $P : \mathcal{B} \to [0, 1]$ with $P(\emptyset) = 0$ and $P(\Omega) = 1$ satisfying the finite superadditivity property that for all $A_1, ..., A_n \in \mathcal{B}$ such that $\bigcup_{i=1}^n A_i \in \mathcal{B}$, *and $(A_i)_{i=1}^n$ are disjoint, we have*

$$\sum_{i=1}^{n} P(A_i) \le P\left(\bigcup_{i=1}^{n} A_i\right) \tag{2.5}$$

and also the countable monotonicity property that

$$P(A) \le \sum_{n=1}^{\infty} P(A_n) \tag{2.6}$$

for all $A, A_1, A_2, ... \in \mathcal{B}$ with $A \subseteq \bigcup_{n=1}^{\infty} A_n$. Then there exists a σ-algebra $\mathcal{C} \supset \mathcal{B}$ and a countably additive probability measure P^* on $\mathcal{C}$ such that $P^*(A) = P(A)$ for all $A \in \mathcal{B}$ (that is, $(\Omega, \mathcal{C}, P^*)$ is a valid probability triple, which agrees with our previous probabilities on $\mathcal{B}$).

Remark 2.3.2

Of course, the conclusions of Theorem 2.3.1 imply that (2.5) must actually hold with *equality*. However, (2.5) need only be verified as an *inequality* to apply Theorem 2.3.1.

However, it is not clear how to even start proving this theorem. Indeed, how could we begin to define $P(A)$ for all A in a σ-algebra? The key is given by *outer measure* P^* defined by

$$P^*(A) = \inf\left\{\sum_{n=1}^{\infty} P(A_n) : A_1, ..., A_n \in \mathcal{B} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \tag{2.7}$$

That is, we define $P^*(A)$ for any subset $A \subseteq \Omega$ to be the infimum of the sums of $P(A_n)$, where $(A_n)_{n=1}^{\infty}$ is any countably collection of the elements of the original semi-algebra $\mathcal{B}$ whose union contains A. In other words, we use the values of $P(A)$ for $A \in \mathcal{B}$ to help us define $P^*(A)$ for any $A \subseteq \Omega$. Of course, we know that P^* will not necessarily be a proper probability measure for all $A \subseteq \Omega$; for example, this is not possible for Uniform($[0, 1]$) by Proposition 1.2.1. However, it is still useful that $P^*(A)$ is at least *defined* for all $A \subseteq \Omega$. We shall eventually show that P^* is indeed a probability measure on some σ-algebra $\mathcal{C}$, and that P^* is an extension of P.

To continue, we note a few simple properties of P^*. Firstly, we clearly have $P^*(\emptyset) = 0$; indeed, we can simply take $A_n = \emptyset$ for all $n \in \mathbb{N}$ in (2.7). Secondly, P^* is *monotone*; indeed, if $A \subseteq B$, then the infimum of (2.7) for $P^*(A)$ includes all choices of $(A_n)_{n=1}^{\infty}$ which work for $P^*(B)$ plus many more besides, so that $P^*(A) \le P^*(B)$. We also have the following lemma.

Lemma 2.3.3

P^* is an extension of P, i.e. $P^*(A) = P(A)$ for all $A \in \mathcal{B}$.

Proof.

Let $A \in \mathcal{B}$ be arbitrary. From (2.6), that $P^*(A) \ge P(A)$. On the other hand, let $A_1 = A$ and $A_n = \emptyset$ for all $n \in \mathbb{N}_{\ge 2}$ in (2.7) shows that $P^*(A) \le P(A)$. ∎

Lemma 2.3.4

P^* is countably subadditive; that is, for all $(B_n)_{n=1}^{\infty}$ of Ω,

$$P^*\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} P^*(B_n)$$

Proof.

Let $(B_n)_{n=1}^{\infty}$ be arbitrary. From (2.7), we see that for any $\varepsilon > 0$, there exists a subsequence $(B_{n_k})_{k=1}^{\infty}$ for each $n \in \mathbb{N}$ with $B_{n_k} \in \mathcal{B}$ such that $B_n \subseteq \bigcup_{k=1}^{\infty} B_{n_k}$ and $\sum_{k=1}^{\infty} P(B_{n_k}) \leq P^*(B_n) + \varepsilon 2^{-n}$. But then the overall collection $(B_{n_k})_{k=1}^{\infty}$ contains $\bigcup_{n=1}^{\infty} B_n$. It follows that $P^*\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n,k=1}^{\infty} P(B_{n_k}) \leq \sum_{n=1}^{\infty} P^*(B_n) + \varepsilon$. Since this is true for any $\varepsilon > 0$, we have that

$$P^*\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} P^*(B_n)$$

as claimed. ■

Now we set

$$\mathcal{C} = \{A \subseteq \Omega : P^*(A \cap E) + P^*(A^c \cap E) = P^*(E), E \subseteq \Omega\} \tag{2.8}$$

That is, $\mathcal{C}$ is the collection of all subsets A with the property that P^* is additive by the union of $A \cap E$ and $A^c \cap E$ for all $E \subseteq \Omega$. Note that it follows by subadditivity that

$$P^*(A \cap E) + P^*(A^c \cap E) \geq P^*(E)$$

so (2.8) is equivalent to

$$\mathcal{C} = \{A \subseteq \Omega : P^*(A \cap E) + P^*(A^c \cap E) \leq P^*(E), E \subseteq \Omega\} \tag{2.9}$$

which is sometimes helpful. Furthermore, P^* is countably additive on $\mathcal{C}$.

Lemma 2.3.5

If $(A_n)_{n=1}^{\infty} \in \mathcal{C}$ are disjoint, then $P^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P^*(A_n)$.

Proof.

If A_1 and A_2 are disjoint with $A_1 \in \mathcal{C}$, then

$$P^*(A_1 \cup A_2) = P^*(A_1 \cap (A_1 \cup A_2)) + P^*(A_1^c \cap (A_1 \cup A_2)) = P^*(A_1) + P^*(A_2)$$

Hence, by induction, the lemma holds for any finite collection of A_n.

Then with countably many disjoint $A_n \in \mathcal{C}$, we see that for any $m \in \mathbb{N}$,

$$\sum_{n=1}^{m} P^*(A_n) = P\left(\bigcup_{n=1}^{m} A_n\right) \leq P^*\left(\bigcup_{n=1}^{\infty} A_n\right)$$

where the inequality follows from monotonicity. Since this holds for any $m \in \mathbb{N}$, we have that

$$\sum_{n=1}^{\infty} P^*(A_n) \leq P^*\left(\bigcup_{n=1}^{\infty} A_n\right)$$

On the other hand, by subadditivity, we have

$$\sum_{n=1}^{\infty} P^*(A_n) \geq P^*\left(\bigcup_{n=1}^{\infty} A_n\right)$$

So the lemma holds for countably many A_n as well. ∎

The plan now is to show that $\mathcal{C}$ is a σ-algebra containing $\mathcal{B}$. We break up the proof into several lemmas.

Lemma 2.3.6

$\mathcal{C}$ is an algebra, i.e. $\Omega \in \mathcal{C}$ and $\mathcal{C}$ is closed under complement and finite intersection (thus, finite union).

Proof.

It follows from (2.8) that $\mathcal{C}$ contains Ω and is closed under complement. For the second statement about finite intersections, suppose $A, B \in \mathcal{C}$. Then for any $E \subseteq \Omega$ and using subadditivity,

$$\begin{aligned}
P^*((A \cap B) \cap E) + P^*((A \cap B)^c \cap E) &= P^*(A \cap B \cap E) \\
&\quad + P^*((A^c \cap B \cap E) \cup (A \cap B^c \cap E) \cup (A^c \cap B^c \cap E)) \\
&\leq P^*(A \cap B \cap E) + P^*(A^c \cap B \cap E) \\
&\quad + P^*(A \cap B^c \cap E) + P^*(A^c \cap B^c \cap E) \\
&= P^*(B \cap E) + P^*(B^c \cap E) && \text{(since } A \in \mathcal{C}) \\
&= P^*(E) && \text{(since } B \in \mathcal{C})
\end{aligned}$$

Therefore, by (2.9), $A \cap B \in \mathcal{C}$. ∎

Lemma 2.3.7

Let $(A_n)_{n=1}^{\infty}$ be disjoint subsets of $\mathcal{C}$. For every $m \in \mathbb{N}$, let

$$B_m = \bigcup_{n=1}^{m} A_n$$

Then for all $m \in \mathbb{N}$ and $E \subseteq \Omega$,

$$P^*(E \cap B_m) = \sum_{n=1}^{m} P^*(E \cap A_n)$$

Proof.

We apply induction on m. Indeed, the statement is trivially true when $m = 1$. Assume that the above statement holds for $m \in \mathbb{N}$. Noting that $B_m \cap B_{m+1} = B_m$ and $B_m^c \cap B_{m+1} = A_{m+1}$, we have $B_m \in \mathcal{C}$ by Lemma 2.3.6 and

$$\begin{aligned} P^*(E \cap B_{m+1}) &= P^*(B_m \cap E \cap B_{m+1}) + P^*(B_m^c \cap E \cap B_{m+1}) \\ &= P^*(E \cap B_m) + P^*(E \cap A_{m+1}) \\ &= \sum_{n=1}^{m+1} P^*(E \cap A_n) \end{aligned}$$

Thereby completing the proof. ■

Lemma 2.3.8

Let $(A_n)_{n=1}^{\infty}$ be a collection of disjoint subsets of $\mathcal{C}$. Then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$.

Proof.

For $m \in \mathbb{N}$, let $B_m = \bigcup_{n=1}^{m} A_n$. Then for any $m \in \mathbb{N}$ and $E \subseteq \Omega$, we have

$$\begin{aligned} P^*(E) &= P^*(E \cap B_m) + P^*(E \cap B_m^c) && \text{(since } B_m \in \mathcal{C}) \\ &= \sum_{n=1}^{m} P^*(E \cap A_n) + P^*(E \cap B_m^c) \\ &\geq \sum_{n=1}^{m} P^*(E \cap A_n) + P^*\left(E \cap \left(\bigcup_{n=1}^{\infty} A_n\right)^c\right) \end{aligned}$$

where the inequality follows from monotonicity as $\left(\bigcup_{n=1}^{\infty} A_n\right)^c \subseteq B_m^c$. this is true for any $m \in \mathbb{N}$, so

$$\begin{aligned} P^*(E) &\geq \sum_{n=1}^{m} P^*(E \cap A_n) + P^*\left(E \cap \left(\bigcup_{n=1}^{\infty} A_n\right)^c\right) \\ &\geq P^*\left(E \cap \left(\bigcup_{n=1}^{\infty} A_n\right)\right) + P^*\left(E \left(\bigcup_{n=1}^{\infty} A_n\right)^c\right) \end{aligned}$$

where the final inequality follows by subadditivity. Thus, by (2.9), we have $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$. ■

Lemma 2.3.9

(i) $\mathcal{C}$ is a σ-algebra.

(ii) $\mathcal{B} \subseteq \mathcal{C}$.

Proof.

To prove the first assertion, in light of Lemma 2.3.6, it suffices to check that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$

whenever $(A_n)_{n=1}^{\infty}$ is a collection in $\mathcal{C}$. Let $D_1 = A_1$, and for $n \geq 2$,

$$D_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i \right)$$

Then clearly, $(D_n)_{n=1}^{\infty}$ are disjoint and $\bigcup_{n=1}^{\infty} D_n = \bigcup_{n=1}^{\infty} A_n$, and with $D_n \in \mathcal{C}$ by Lemma 2.3.6, $\bigcup_{n=1}^{\infty} D_n \in \mathcal{C}$, and so $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$.

To prove the second assertion, let $A \in \mathcal{B}$. Then since $\mathcal{B}$ is a semi-algebra, we can write

$$A^c = \bigsqcup_{i=1}^{m} B_m$$

for some disjoint $B_1, ..., B_m \in \mathcal{B}$. Also, for any $E \subseteq \Omega$ and $\varepsilon > 0$, by definition (2.7), there exists a collection $(A_n)_{n=1}^{\infty}$ of $\mathcal{B}$ with $E \subseteq \bigcup_{n=1}^{\infty} A_n$ and $\sum_{n=1}^{\infty} P(A_n) \leq P^*(E) + \varepsilon$. Then by monotonicity, subadditivity, and since $P^* = P$ on $\mathcal{B}$, we have that

$$\begin{aligned}
P^*(E \cap A) + P^*(E \cap A^c) &\leq P\left(\left(\bigcup_{n=1}^{\infty} A_n \right) \cap A \right) + P^*\left(\left(\bigcup_{n=1}^{\infty} A_n \right) \cap A^c \right) \\
&= P^*\left(\bigcup_{n=1}^{\infty} (A_n \cap A) \right) + P^*\left(\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{m} (A_n \cap B_i) \right) \\
&= \sum_{n=1}^{\infty} P(A_n \cap A) + \sum_{n=1}^{\infty} \sum_{i=1}^{m} P^*(A_n \cap B_i) \\
&= \sum_{n=1}^{\infty} \left(P(A_n \cap A) + \sum_{i=1}^{m} P(A_n \cap B_i) \right) \\
&\leq \sum_{n=1}^{\infty} P(A_n) \\
&\leq P^*(E) + \varepsilon
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we have

$$P^*(E \cap A) + P^*(E \cap A^c) \leq P^*(E)$$

for all $E \subseteq \Omega$. Hence, by (2.9), $A \in \mathcal{C}$. This holds for any $A \in \mathcal{B}$, so $\mathcal{B} \subseteq \mathcal{C}$. ∎

With all of the lemmas at our disposal, proving Theorem 2.3.1 is very easy.

Proof.

(of Theorem 2.3.1) Lemmas 2.3.3, 2.3.5, and 2.3.9 together show that $\mathcal{C}$ is a σ-algebra containing $\mathcal{B}$, that P^* is a probability measure on $\mathcal{C}$, and that P^* is an extension of P. ∎

Exercise 2.3.10

Prove that the extension $(\Omega, \mathcal{C}, P^*)$ constructed in the proof of Theorem 2.3.1 is *complete*. That is, if

(i) If $A \in \mathcal{C}$ with $P^*(A) = 0$.

(ii) $B \subseteq A$,

then $B \in \mathcal{C}$.

2.4 Constructing the Uniform$[0,1]$ Distribution*

Theorem 2.3.1 allows us to automatically construct valid probability spaces which take particular values on particular sets. We now use this to construct Uniform$[0,1]$, the uniform distribution on $[0,1]$. We begin by letting $\Omega = [0,1]$, and let

$$\mathcal{B} = \{\text{all intervals contained in } [0,1]\} \tag{2.10}$$

where the "intervals" is understood to include all the open, closed, half-open, and singletons contained in $[0,1]$, and also the emptyset $\emptyset$. Then $\mathcal{B}$ is a semi-algebra by Exercise 2.2.4.

For $I \in \mathcal{B}$, we let $P(I)$ to denote the length of I. Thus, $P(\emptyset) = 0$ and $P(\Omega) = 1$. We now proceed to verify (2.5) and (2.6).

Proposition 2.4.1

The above definition of $\mathcal{B}$ and P satisfies (2.5) with equality.

Proof.

Let $I_1, ..., I_n$ be disjoint intervals in $[0,1]$ where the union is some interval I_0. For $0 \le i \le n$, write a_i for the left endpoint of I_i and b_i for the right endpoint of I_i. The assumptions imply that by reordering, we can ensure that $a_0 = a_1 \le b_1 = a_2 \le \cdots \le b_k = b_0$. Then

$$\sum_{i=1}^{n} P(I_i) = \sum_{i=1}^{n} (b_i - a_i) = b_k - a_1 = b_0 - a_0 = P(I_0)$$

■

The verification of (2.6) for such a $\mathcal{B}$ and P is a bit more involved.

Exercise 2.4.2

(a) Prove that if $I_1, ..., I_n$ is a collection of intervals, and if $\bigcup_{i=1}^n I_i \supset I$ for some interval I, then

$$\sum_{i=1}^{n} P(I_i) \ge I$$

Hint: Use the ideas of the proof of Proposition 2.4.1.

(b) Prove that $(I_n)_{n=1}^{\infty}$ is a countable collection of open intervals, and if $\bigcup_{n=1}^{\infty} \supset I$ for some closed interval I, then

$$\sum_{n=1}^{\infty} P(I_n) \geq P(I)$$

Hint: You may use the Heine-Borel Theorem, which says that if a collection of open intervals contain a closed interval, then some finite subcollection of the open intervals also contains the closed interval.

(c) Verify (2.6), i.e. prove that if $(I_n)_{n=1}^{\infty}$ is any countable collection of intervals and if $\bigcup_{n=1}^{\infty} I_n \supset I$ for any interval I, then $\sum_{n=1}^{\infty} P(I_n) \geq P(I)$.

Hint: Extend the interval I_n by $\varepsilon/2^n$ at the end, and decrease I by ε at each end, while making I_n open and I closed. Then use (b).

In light of Proposition 2.4.1 and Exercise 2.4.2, we can apply Theorem 2.3.1 to conclude the following.

Theorem 2.4.3

There exists a probability space $(\Omega, \mathcal{A}, P^*)$ such that $\Omega = [0,1]$, $\mathcal{A}$ contains all intervals in $[0,1]$, and for any $I \subseteq [0,1]$, $P^*(I)$ is the length of I.

This probability triple is called either the uniform distribution on $[0,1]$, or the *Lebesgue measure on* $[0,1]$. Depending on the context, sometimes we write the probability measure P^* as P or λ.

Remark 2.4.4

Let $\sigma(\mathcal{B})$ be the *σ-algebra generated by* $\mathcal{B}$, i.e. the smallest σ-algebra containing $\mathcal{B}$. The collection $\sigma(\mathcal{B})$ is called the *Borel σ-algebra* of subsets of $[0,1]$, and the elements of $\sigma(\mathcal{B})$ are called *Borel sets*. Clearly, we have $\sigma(\mathcal{B}) \supset \mathcal{A}$. In this case, it can be shown that $\mathcal{A}$ is much larger than $\sigma(\mathcal{B})$; it even has larger cardinality. Furthermore, it turns out that Lebesgue measure restricted to $\sigma(\mathcal{B})$ is not complete, though $\mathcal{A}$ is complete by Exercise 2.3.10. In addition to the Borel subsets of $[0,1]$, we also have occasion to refer the *Borel σ-algebra of subsets of* $\mathbb{R}$, defined to be the smallest σ-algebra of subsets of $\mathbb{R}$ which includes all intervals.

Exercise 2.4.5

Let $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$. Prove that $\sigma(\mathcal{A}) = \sigma(\mathcal{B})$, i.e. the smallest σ-algebra of subsets of $\mathbb{R}$ which contains $\mathcal{A}$ is equal to the Borel σ-algebra of subsets of $\mathbb{R}$.

Hint: Does $\sigma(\mathcal{A})$ contain all intervals?

Writing λ for Lebesgue measure on $[0,1]$, we know that $\lambda(\{x\}) = 0$ for any singleton $\{x\}$. It follows by countable additivity that $\lambda(A) = 0$ for any set A which is countable. This includes the rational numbers $\mathbb{Q}$, the integer roots of the rational numbers, the algebraic numbers, etc. That is, if X is uniformly distributed on $[0,1]$, then $P(X \in \mathbb{Q}) = 0$ and $P(X^n \in \mathbb{Q} \text{ for some } n \in \mathbb{N}) = 0$ and $P(X \in \mathbb{A}) = 0$, where $\mathbb{A}$ denotes the set of all algebraic numbers, and so on.

There also exist uncountable sets which have Lebesgue measure 0. The simplest example is the *Cantor set* K, defined as follows:

Figure 2.1: Constructing the Cantor Set K

We begin with the interval $[0,1]$. We remove the open interval consisting of the middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$. We then remove the open middle third of each of the two pieces, i.e. we remove $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$. Then remove the four open middle thirds $\left(\frac{1}{27}, \frac{2}{27}\right)$, $\left(\frac{7}{27}, \frac{8}{27}\right)$, $\left(\frac{19}{27}, \frac{20}{27}\right)$, and $\left(\frac{25}{27}, \frac{26}{27}\right)$ of the remaining pieces. We continue inductively, at the nth stage removing the 2^{n-1} middle thirds of all remaining subintervals, each of length $\frac{1}{3^n}$. The Cantor set K is defined to be everything that is left over, after we have removed the middle thirds.

Now, the complement of the Cantor set has Lebesgue measure

$$\lambda(K^c) = \frac{1}{3} + 2 \cdot 194 \cdot \frac{1}{7} + \cdots = \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1$$

Therefore, $\lambda(K) = 0$.

On the other hand, K is uncountable. Indeed, for each $x \in K$, let $d_n(x) = 0$ or 1, depending on whether at the nth stage of the construction of K, x was to the left or right of the nearest open interval removed. Then define $f : K \to [0,1]$ by

$$f(x) = \sum_{n=1}^{\infty} d_n(x) 2^{-n}$$

It is easily checked that $f(K) = [0,1]$, i.e. f maps K onto $[0,1]$. Since $[0,1]$ is uncountable, then K must also be uncountable.

Remark 2.4.6

The Cantor set is also equal to the set of all numbers in $[0,1]$ which have a base-3 expansion that does not contain the digit 1. That is,

$$K = \left\{ \sum_{n=1}^{\infty} c_n 3^n : c_n \in \{0,2\} \right\}$$

Exercise 2.4.7

(a) Prove that K and K^c are in $\sigma(\mathcal{B})$, where $\sigma(\mathcal{B})$ are the Borel subsets of $[0,1]$.

(b) Prove that K and K^c are in $\mathcal{A}$, where $\mathcal{A}$ is the σ-algebra defined in Theorem 2.4.3.

On the other hand, from Proposition 1.2.1, we know that

Proposition 2.4.8

For the probability space $(\Omega, \mathcal{A}, P^*)$ *of Theorem 2.4.3 corresponding to Lebesgue measure* $[0,1]$, *there exists a subset* $H \subseteq \Omega$ *such that* $H \notin \mathcal{A}$.

2.5 Corollaries of the Extension Theorem*

The Extension Theorem (2.3.1) will be our main tool for proving the existence of complicated probability spaces. While (2.5) is generally easy to verify, (2.6) can be more challenging. Thus, we present some alternative formulations here.

Corollary 2.5.1

Let $\mathcal{A}$ *be a semi-algebra of subsets of* Ω. *Let* $P : \mathcal{A} \to [0,1]$ *with* $P(\emptyset) = 0$ *and* $P(\Omega) = 1$ *satisfying (2.5) and the* monotonicity property *on* $\mathcal{A}$ *that if* $A, B \in \mathcal{A}$ *are such that* $A \subseteq B$, *then*

$$P(A) \leq P(B) \tag{2.11}$$

and also the countable subadditivity *on* $\mathcal{A}$ *that*

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} P(B_n) \tag{2.12}$$

for $(B_n)_{n=1}^{\infty}$ *in* $\mathcal{A}$ *such that* $\bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. *Then there exists a* σ*-algebra* $\mathcal{B} \supset \mathcal{A}$ *and a countably additive probability measure* P^* *on* $\mathcal{B}$ *such that* $P^*(A) = P(A)$ *for all* $A \in \mathcal{A}$.

Proof.

In light of Theorem 2.3.1, we need only verify (2.6). Indeed, let A and $(A_n)_{n=1}^{\infty}$ be in $\mathcal{A}$ with $A \subseteq \bigcup_{n=1}^{\infty}$. Let $B_n = A \cap A_n$. Then since $A \subseteq \bigcup_{n=1}^{\infty} A_n$, we have $A = \bigcup_{n=1}^{\infty}(A \cap A_n) = \bigcup_{n=1}^{\infty} B_n$, and so (2.11) and (2.12) give that

$$P(A) = P\left(\sum_{n=1}^{\infty} B_n\right) \leq \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$$

This completes the proof. ∎

Another version assumes countable additivity on $\mathcal{A}$.

Corollary 2.5.2

Let $\mathcal{A}$ *be a semi-algebra of subsets of* Ω. *Let* $P : \mathcal{A} \to [0,1]$ *with* $P(\Omega) = 1$ *satisfying the* *countable additivity property that*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \tag{2.13}$$

for all $(A_n)_{n=1}^{\infty}$ *of* $\mathcal{A}$ *disjoint and* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. *Then there exists a* σ*-algebra* $\mathcal{B} \supset \mathcal{A}$ *and a countably additive probability measure* P^* *on* $\mathcal{B}$ *such that* $P^*(A) = P(A)$ *for all* $A \in \mathcal{A}$.

Proof.

Note that (2.13) automatically implies (2.5) with equality and that $P(\emptyset) = 0$. Hence, in light of Corollary 2.5.1, we need to verify (2.11) and (2.12).

For (2.11), let $A, B \in \mathcal{A}$ with $A \subseteq b$. Since $\mathcal{A}$ is a semi-algebra, then we can write

$$A^c = \bigsqcup_{i=1}^{k} A_i$$

for some disjoint $A_1, ..., A_k \in \mathcal{A}$. Then using (2.13), we have

$$P(B) = P(A) + \sum_{i=1}^{k} P(B \cap A_i) \geq P(A)$$

For (2.12), let $(B_n)_{n=1}^{\infty}$ be a collection in $\mathcal{A}$ with $\bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. Let $C_1 = B_1$ and for $n \geq 2$,

$$C_n = B_n \cap \bigcap_{i=1}^{n-1} B_i^c$$

Then $(C_n)_{n=1}^{\infty}$ are disjoint with $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} C_n$. Furthermore, since $\mathcal{A}$ is a semi-algebra, each C_n can be written as a finite disjoint union of elements of $\mathcal{A}$, say

$$C_n = \bigcup_{i=1}^{k_n} J_{n_i}$$

It then follows from (2.11) and (2.5) that

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} B_n\right) &= P\left(\bigcup_{n=1}^{\infty} C_n\right) \\
&= P\left(\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{k_n} J_{n_i}\right) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{k_n} P(J_{n_i}) \\
&= \sum_{n=1}^{\infty} P(C_n)
\end{aligned}$$

$$\leq \sum_{n=1}^{\infty} P(B_n)$$

This completes the proof. ■

Exercise 2.5.3

Suppose P satisfies (2.13) for finite collections $(C_n)_{n=1}^{\infty}$. Suppose further that whenever $(A_n)_{n=1}^{\infty}$ are in $\mathcal{A}$ such that $A_{n+1} \subseteq A_n$ and $\bigcup_{n=1}^{\infty} A_n = \emptyset$, prove that $\lim_{n\to\infty} P(A_n) = 0$. Prove that P also satisfies (2.13) for countable collections $(C_n)_{n=1}^{\infty}$.

The extension of Theorem 2.3.1 also has a uniqueness property.

Proposition 2.5.4

Let $\mathcal{A}$, P, P^*, and $\mathcal{B}$ be as in Theorem 2.3.1. Let $\mathcal{F}$ be any σ-algebra with $\mathcal{A} \subseteq \mathcal{F} \subseteq \mathcal{B}$. Let Q be any probability measure on $\mathcal{F}$ such that $Q(A) = P(A)$ for all $A \in \mathcal{A}$. Then $Q(A) = P^*(A)$ for all $A \in \mathcal{F}$.

Proof.

Let $A \in \mathcal{F}$ be arbitrary. Then by definition of the outer measure,

$$\begin{aligned}
P^*(A) &= \inf\left\{\sum_{n=1}^{\infty} P(A_i) : (A_n)_{n=1}^{\infty} \text{ is a countable collection of } \mathcal{A} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \\
&= \inf\left\{\sum_{n=1}^{\infty} Q(A_i) : (A_n)_{n=1}^{\infty} \text{ is a countable collection of } \mathcal{A} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \\
&\qquad \text{(since } Q = P \text{ on } \mathcal{A}) \\
&\geq \inf\left\{Q\left(\bigcup_{n=1}^{\infty}\right) : (A_n)_{n=1}^{\infty} \text{ is a countable collection of } \mathcal{A} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \\
&\qquad \text{(by countable subadditivity)} \\
&\geq \inf\left\{Q(A) : (A_n)_{n=1}^{\infty} \text{ is a countable collection of } \mathcal{A} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \\
&\qquad \text{(by monotonicity)} \\
&= Q(A)
\end{aligned}$$

i.e. $P^*(A) \geq Q(A)$. Similarly, $P^*(A^c) \geq Q(A^c)$ and so $1 - P^*(A) \geq 1 - Q(A)$ implies $P^*(A) \leq Q(A)$. Thus, $P^*(A) = Q(A)$. This completes the proof. ■

Corollary 2.5.5

Let $\mathcal{A}$ be a semi-algebra of subsets of Ω, and let $\mathcal{F} = \sigma(\mathcal{A})$ be the generated σ-algebra. Let P and $Q/$ be two probability distributions defined on $\mathcal{F}$. Suppose that $P(A) = Q(A)$ for all

$A \in \mathcal{A}$. Then $P = Q$.

Proof.

Since P and Q are probability measures, they both satisfy (2.5) and (2.6). Hence, by Proposition 2.5.4, each P and Q is equal to P^*. ■

One useful special case of Corollary 2.5.5 is the following.

Corollary 2.5.6

Let P and Q be two probability distributions defined on the collection $\mathfrak{B}(\mathbb{R})$ of Borel subsets of $\mathbb{R}$. Suppose $P((-\infty, x]) = Q((-\infty, x])$ for all $x \in \mathbb{R}$. Then $P(A) = Q(A)$ for all $A \in \mathfrak{B}(\mathbb{R})$.

Proof.

Since $P((y, \infty)) = 1 - P((-\infty, y])$ and $P((x, y]) = 1 - P((-\infty, x]) - P((y, \infty))$, and similarly, for Q, it follows that P and Q agree on

$$\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\} \cup \{(y, \infty) : y \in \mathbb{R}\} \cup \{(x, y] : x < y \in \mathbb{R}\} \cup \{\emptyset, \mathbb{R}\}$$

But since $\mathcal{A}$ is a semi-algebra, and it follows from Exercise 2.4.5 that $\sigma(\mathcal{A}) = \mathfrak{B}(\mathbb{R})$. Hence, the result follows from Corollary 2.5.5. ■

2.6 Coin Tossing and Other Measures

Now that we have Theorem 2.3.1 to help us, we can construct other probability spaces as well.

For example, of frequent mention in probability theory is (independent fair) coin tossing. To model the flipping of n coins, we can simply take

$$\Omega = \{(x_i)_{i=1}^n : x_i = H \text{ or } x_i = T\}$$

where H stands for heads and T stands for tails. Let $\mathcal{F} = \mathcal{P}(\Omega)$ be the collection of all subsets of Ω and define P by

$$P(A) = \frac{|A|}{2^n}$$

for $A \subseteq \mathcal{F}$. This is another example of a discrete probability space; and we know from Theorem 2.2.2 that these spaces present no difficulties.

But suppose now that we wish to model the flipping of a countable *infinite* number of coins. In this case, we let

$$\Omega = \{H, T\}^{\mathbb{N}} = \{(x_n)_{n=1}^\infty : x_n = H \text{ or } x_n = T\}$$

be the collection of all binary sequences. But what about $\mathcal{F}$ and P? For each $n \in \mathbb{N}$ and $e_1, ..., e_n \in \{H, T\}$, let us define subsets $A_{x_1,...,x_n} \subseteq \Omega$ by

$$A_{e_1,...,e_n} = \{(x_n)_{n=1}^\infty \in \Omega : x_i = e_i \text{ for all } 1 \leq i \leq n\}$$

Thus, A_T is the event that the first coin comes up tails, $A_{H,H}$ is the event that the first two coins both come up heads; and $A_{H,T,H}$ is the event that the first and third coins are heads while the second coin is tails. Then we clearly want

$$P(A_{e_1,\dots,e_n}) = \frac{1}{2^n}$$

for each $A_{e_1,\dots,e_n}$. Hence, if we set

$$\mathcal{A} = \{A_{e_1,\dots,e_n} : n \in \mathbb{N}, e_1, \dots, e_n \in \{H, T\}\} \cup \{\emptyset, \Omega\}$$

then we know how to define $P(A)$ for each $A \in \mathcal{A}$. To apply the Extension Theorem, we need to verify that certain conditions are satisfied.

Verifying (**??**) for countable collections requires a bit of topology.

Lemma 2.6.1

The above $\mathcal{A}$ and P for infinite coin tossing satisfy (2.13).

Proof.

Let $(\{H, T\}^{\mathbb{N}}, \mathcal{T})$ be the topological space such that $\{H, T\}$ has $\mathcal{T}$ being the discrete topology, and give $\Omega = \{H, T\}^{\mathbb{N}}$ the corresponding product topology. Then Ω is a product of compact sets $\{H, T\}$ and hence is itself compact by Tikhonov's Theorem.

Furthermore, each element in $\mathcal{A}$ is a closed subset of Ω, since its complement is open in the product topology. Hence, each A_n is a closed subset of a compact space, and therefore compact.

The *finite intersection property* of compact sets then tells us that there exists an $N \in \mathbb{N}$ such that $A_n = \emptyset$ for all $n > N$. In particular, $\lim_{n\to\infty} P(A_n) = 0$. This completes the proof. ∎

Now that the conditions have been verified, then by Corollary 2.5.2 that the probabilities for the special sets $A_{e_1,\dots,e_n} \in \mathcal{A}$ can automatically be extended to a σ-algebra $\mathcal{B}$ containing $\mathcal{A}$. This will be our probability space for infinite fair coin tossing.

Example 2.6.2

Let $H_n = \{(x_n)_{n=1}^{\infty} : x_n = H\}$ be the event that the nth coin comes up heads. We certainly would hope that $H_n \in \mathcal{B}$ with $P(H_n) = \frac{1}{2}$. Indeed, note that

$$H_n = \bigsqcup_{x_1,\dots,x_{n-1} \in \{H,T\}} A_{x_1,\dots,x_{n-1},H}$$

the union being disjoint. Hence, since $\mathcal{B}$ is closed under countable unions, we have $H_n \in \mathcal{B}$. Then by countable additivity,

$$P(H_n) = \sum_{x_1,\dots,x_{n-1} \in \{H,T\}} P(A_{x_1,\dots,x_{n-1},H})$$

$$
\begin{aligned}
&= \sum_{x_1,\ldots,x_{n-1}\in\{H,T\}} \frac{1}{2^n} \\
&= \frac{2^{n-1}}{2^n} \\
&= \frac{1}{2}
\end{aligned}
$$

Remark 2.6.3

If we identify an element $x \in [0,1]$ by its binary expansion $(x_n)_{n=1}^{\infty}$ so that

$$x = \sum_{n=1}^{\infty} \frac{x_n}{2^n}$$

then we see that in fact infinite fair coin tossing may be viewed as being "essentially" the same thing as the Lebesgue measure on $[0,1]$.

For the remainder of the section, we will define product measures on probability spaces.

Definition 2.6.4

Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega, \mathcal{F}_2, P_2)$ be two probability spaces. We can define their *product measure* P on the Cartesian product set

$$\Omega_1 \times \Omega_2 = \{(x_1, x_2) : x_1 \in \Omega_1, x_2 \in \Omega_2\}$$

We set

$$\mathcal{F}_{12} = \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$$

and define $P(A \times B) = P_1(A)P_2(B)$ for $A \times B \in \mathcal{F}_{12}$. The elements of $\mathcal{F}_{12}$ are called *measurable rectangles.*

Exercise 2.6.5

Show that the above $\mathcal{F}_{12}$ is a semi-algebra and that $\emptyset, \Omega \in \mathcal{F}_{12}$ with $P(\emptyset) = 0$ and $P(\Omega) = 1$.

Chapter 3

Further Probabilistic Foundations

Now that we understand probability spaces well, we discuss some additional essential ingredients of probability theory. Throughout this chapter, we shall assume that there is an underlying probability space $(\Omega, \mathcal{F}, P)$ with respect to which all further probability objects are defined. This assumption shall be so universal that we will often not even mention it.

3.1 Random Variables

If we think of a sample space Ω as the set of all possible random outcomes of some experiment, then a *random variable* assigns a numerical value to each of these outcomes. More formally,

Definition 3.1.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \tag{3.1}$$

for $x \in \mathbb{R}$.

Equation (3.1) is a technical requirement, and states that the function X must be *measurable*. It can also be written as $\{X \leq x\} \in \mathcal{F}$ or $X^{-1}[(-\infty, x]] \in \mathcal{F}$ for all $x \in \mathbb{R}$. Since complements and unions and intersections are preserved under inverse images, it follows that (3.1) is equivalent to saying that $X^{-1}[B] \in \mathcal{F}$ for every Borel set B. That is, the set $X^{-1}[B]$, also written $\{X \in B\}$ is indeed an event. So for any Borel set B, it makes sense to talk about $P(X \in B)$, the probability that X lies in B.

Example 3.1.2

Suppose that $(\Omega, \mathcal{F}, P)$ is the Lebesgue measure on $[0, 1]$. Then we might define some random variables X, Y and Z by $X(\omega) = \omega$, $Y(\omega) = 2\omega$, and $Z(\omega) = 3\omega + 4$. Then we have $Y = 2X$

and $Z = 3X + 4 = \frac{3}{2}Y + 4$. Also,

$$P\left(Y \leq \frac{1}{3}\right) = P\left(\left\{\omega : Y(\omega) < \frac{1}{3}\right\}\right) = P\left(\left\{\omega : 2\omega < \frac{1}{3}\right\}\right) = P\left(\left[0, \frac{1}{6}\right]\right) = \frac{1}{6}$$

Exercise 3.1.3

For Example 3.1.2, compute $P(Z > a)$ and $P(X < a, Y < b)$ as functions of $a, b \in \mathbb{R}$.

Now not all functions from Ω to $\mathbb{R}$ are random variables. For example, let $(\Omega, \mathcal{F}, P)$ be the Lebesgue measure on $[0, 1]$ and let $H \subseteq \Omega$ be the non-measurable set of Proposition 2.4.8. Define $X : \Omega \to \mathbb{R}$ by

$$X = \mathbf{1}_{H^c}(x) = \begin{cases} 0 & \text{if } x \in H \\ 1 & \text{if } x \notin H \end{cases}$$

Then $\left\{\omega \in \Omega : X(\omega) \leq \frac{1}{2}\right\} = H^c \notin \mathcal{F}$, so X is not a random variable.

On the other hand, the following proposition shows that (3.1) is preserved under usual arithmetic and limits. In practice, this means that if functions from Ω to $\mathbb{R}$ are constructed in "usual" ways, then (3.1) will be satisfied, so the functions will indeed be random variables.

Proposition 3.1.4

(i) If $X = \mathbf{1}_A$ is the indicator for some $A \in \mathcal{F}$, then X is a random variable.

(ii) If X and Y are random variables and $c \in \mathbb{R}$, then $X + c$, cX, X^2, $X + Y$ and XY are all random variables.

(iii) If $(X_n)_{n=1}^{\infty}$ is a collection of random variables such that $\lim_{n\to\infty} X_n(\omega)$ exists for each $\omega \in \Omega$ and $X(\omega) = \lim_{n\to\infty} X_n(\omega)$, then X is also a random variable.

Proof.

To see that (i) is true, note that for any $A \in \mathcal{F}$, then

$$\mathbf{1}_A^{-1}[B] = \begin{cases} \Omega & \text{if } 0, 1 \in B \\ A^c & \text{if } 1 \notin B,\, 0 \in B \\ A & \text{if } 1 \in B,\, 0 \notin B \\ \emptyset & \text{if } 0, 1 \notin B \end{cases}$$

Therefore, $\mathbf{1}_A^{-1}[B] \in \mathcal{F}$.

To see that (ii) is true, note that the first two assertions are immediate. The third one follows since for $y \geq 0$, $\{X^2 \leq y\} = \{X \in [-\sqrt{y}, \sqrt{y}]\} \in \mathcal{F}$. For the fourth, note that by the density of the rationals, if $r \in (X, x - Y)$, we have

$$\{X + Y < x\} = \bigcup_{r\in\mathbb{Q}} (\{X < r\} \cap \{Y < x - r\}) \in \mathcal{F}$$

Finally, the fifth assertion then follows since

$$XY = \frac{1}{2}[(X + Y)^2 - X^2 - Y^2]$$

To see that (iii) is true, for $x \in \mathbb{R}$, we have that

$$\{X \leq x\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{X_k \leq x + \frac{1}{m}\right\}$$

But X_k is a random variable, so $\left\{X_k \leq x + \frac{1}{m}\right\} \in \mathcal{F}$. Then since $\mathcal{F}$ is a σ-algebra, it follows that $\{X \leq x\} \in \mathcal{F}$.

Definition 3.1.5

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *Borel measurable* if and only if $f^{-1}[A] \in \mathfrak{B}(\mathbb{R})$ for any $A \in \mathfrak{B}(\mathbb{R})$.

Definition 3.1.6

Let X be a random variable and let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then we define the *composition* of X with f by $f(X(\omega))$ for each $\omega \in \Omega$. Then (3.1) is satisfied since for $B \in \mathfrak{B}(\mathbb{R})$, $\{f(X) \in B\} = \{X \in f^{-1}[B]\} \in \mathcal{F}$.

Proposition 3.1.7

Every continuous function f is Borel measurable.

Proof.

A basic result from point-set topology says that if f is continuous, then $f^{-1}[V]$ is open subset of $\mathbb{R}$ if V is open. In particular, $f^{-1}[(x, \infty)]$ is open, so $f^{-1}[(x, \infty)] \in \mathfrak{B}(\mathbb{R})$, so $f^{-1}[(-\infty, x]] \in \mathfrak{B}(\mathbb{R})$.

For example, if $f(x) = x^n$ for $n \in \mathbb{N}$, then f is Borel-measurable. Hence, if X is a random variable, then so is X^n for all $n \in \mathbb{N}$.

Remark 3.1.8

In probability theory, the underlying probability space $(\Omega, \mathcal{F}, P)$ is usually *complete*. In that case, if X is a random variable and $Y : \Omega \to \mathbb{R}$ such that $P(X = Y) = 1$, then Y is also a random variable.

Remark 3.1.9

In Definition 3.1.1, we assume that X is a real-valued random variable. More generally, we could consider a random variable which mapped Ω to an arbitrary measurable space, i.e. to some secondary nonempty Ω' with its own collection $\mathcal{F}'$ of measurable subsets. We would then have $X : \Omega \to \Omega'$ with condition (3.1) replaced by the condition that $X^{-1}[A'] \in \mathcal{F}$ if $A' \in \mathcal{F}$.

3.2 Independence

Definition 3.2.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(i) Two events $A, B \in \mathcal{F}$ are said to be *independent* if

$$P(A \cap B) = P(A)P(B)$$

(ii) A collection $(A_i)_{i \in I} \in \mathcal{F}$ is said to be *independent* if for any finite $F \subseteq I$,

$$P\left(\bigcap_{i \in F} A_i\right) = \prod_{i \in F} A_i$$

(iii) Two random variables X and Y are said to be *independent* if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all $X, Y \in \mathfrak{B}(\mathbb{R})$.

(iv) A collection of random variables $(X_i)_{i \in I}$ are said to be *independent* if for any finite $F \subseteq I$,

$$P\left(X_i \in A_i, \forall i \in F\right) = \prod_{i \in F} P(X_i \in A_i)$$

In most cases, instead of talking about an arbitrary collection of events or random variables, it is useful to consider finite collections of events or random variables. Indeed,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i) \tag{3.2}$$

Proposition 3.2.2

Let X and Y be independent random variables on $(\Omega, \mathcal{F}, P)$ and let $f, g : \mathbb{R} \to \mathbb{R}$ be Borel measurable functions. Then the random variables $f(X)$ and $g(Y)$ are independent.

Proof.

For Borel $V_1, V_2 \subseteq \mathbb{R}$, we compute that

$$\begin{aligned} P(f(X) \in V_1, g(Y) \in V_2) &= P(X \in f^{-1}(V_1), Y \in g^{-1}(V_2)) \\ &= P(X \in f^{-1}(V_1))P(Y \in g^{-1}(V_2)) \\ &\qquad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= P(f(X) \in V_1)P(g(Y) \in V_2) \end{aligned}$$

Therefore, $f(X)$ and $g(Y)$ are independent. ∎

Proposition 3.2.3

Let X and Y be random variables on $(\Omega, \mathcal{F}, P)$. The following are equivalent.

(a) X and Y are independent.

(b) $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all $x, y \in \mathbb{R}$.

Proof.

(a) ⇒ (b) follows immediately by definition.

(b) ⇒ (a) Let $x \in \mathbb{R}$ be such that $P(X \leq x) > 0$ and define the measure Q on the Borel subsets of $\mathbb{R}$ by

$$Q(S) = \frac{P(X \leq x, Y \in S)}{P(X \leq x)}$$

By assumption, $Q((-\infty, y]) = P(Y \leq y)$ for all $y \in \mathbb{R}$. It follows from Corollary 2.5.6 that $Q(S) = P(Y \in S)$ for all $S \subseteq \mathbb{R}$, i.e.

$$P(X \leq x, Y \in S) = P(X \leq x)P(Y \in S)$$

Then for any fixed $S \subseteq \mathbb{R}$, let

$$R(T) = \frac{P(X \in T, Y \in S)}{P(X \in T)}$$

Then by the above, we have $R((-\infty, x]) = P(X \leq x)$ for each $x \in \mathbb{R}$. Then by Corollary 2.5.6, $R(T) = P(X \in T)$ for all Borel $T \subseteq \mathbb{R}$, i.e. that $P(X \in T, Y \in S) = P(X \in T)P(Y \in S)$ for all Borel $S, T \subseteq \mathbb{R}$. Hence, X and Y are independent. ∎

3.3 Continuity of Probabilities

Given a probability space $(\Omega, \mathcal{F}, P)$ and event $A \in \mathcal{F}$ and a collection $(A_n)_{n=1}^{\infty}$ of $\mathcal{F}$, we denote $A_n \nearrow A$ to denote that A_n is an increasing collection of events in $\mathcal{F}$, i.e. $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ and also $\bigcup_{n=1}^{\infty} A_n = A$. In other words, the events A_n increase to A. Similarly, we write $(A_n) \searrow A$ to denote that $(A_n^c) \nearrow A^c$ or equivalently, $A_1 \supset A_2 \supset A_3 \supset \cdots$ and $\bigcap_{n=1}^{\infty} A_n = A$. In other words, A_n decrease to A.

Proposition 3.3.1: Monotone Convergence Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ be a sequence of events.

(i) If $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right)$$

(ii) If $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

Proof.

To see that (i) holds, let $A_0 = \emptyset$ for conventional simplicity, and for each $n \in \mathbb{N}$, define

$$B_n = A_n \setminus A_{n-1}$$

then $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ such that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ and $A_n = \bigcup_{i=1}^{n} B_i$ for each $n \in \mathbb{N}$. Hence,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty}\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} P(A_n)$$

proving (i).

To see that (ii) holds, let $B_n = A_1 \setminus A_n$ for each $n \in \mathbb{N}$. Then $\{B_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ and $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$. Then since

$$\bigcup_{n=1}^{\infty} B_n = A_1 \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)$$

it follows from (i) that

$$P\left(A_1 \setminus \left(\bigcap_{n=1}^{\infty}\right) A_n\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty} P(A_1 \setminus A_n)$$

Since $P(A_1 \setminus E) = P(A_1) - P(E)$ for each $E \in \mathcal{F}$ and $E \subseteq A_1$,

$$\begin{aligned} P(A_1) - P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(A_1 \setminus \left(\bigcap_{n=1}^{\infty} A_n\right)\right) \\ &= \lim_{n\to\infty} P(A_1 \setminus A_n) \\ &= P(A_1) - \lim_{n\to\infty} P(A_n) \end{aligned}$$

in which by subtracting $P(A_1)$ on both sides, we obtain the desired equality, proving (ii). ■

If the $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ are not nested, then we may not have $\lim_{n\to\infty}(A_n) = P(A)$, where A is either $\bigcup_{n=1}^{\infty} A_n$ or $\bigcap_{n=1}^{\infty} A_n$.

Example 3.3.2

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ be a sequence of events defined by

$$A_n = \begin{cases} \Omega & \text{if } n \text{ is odd} \\ \emptyset & \text{if } n \text{ is even} \end{cases}$$

Then clearly, $P(A_n)$ alternates between 0 and 1, so that $\lim_{n\to\infty} P(A_n)$ does not exist.

3.4 Limit Events

Definition 3.4.1

Let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$.

(i) We define

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

in which, we say that the events A_n occur *infinitely often*, denoted as A_n i.o.

(ii) We define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

in which, we say that the events A_n occur *almost always*, denoted as A_n a.a.

Intuitively, $\limsup_{n\to\infty} A_n$ is the event that infinitely many of the events A_n occur, while $\liminf_{n\to\infty} A_n$ is the event that all but a finite number of events A_n occur.

Example 3.4.2

Let $(\Omega, \mathcal{F}, P)$ be the probability space of infinite fair coin tossing, and H_n is the event that the nth coin is heads. Then $\limsup_{n\to\infty} H_n$ is the event that there are infinitely many heads, and $\limsup_{n\to\infty} H_n$ is the event that all but a finite number of the coins were heads, i.e. there are finitely many tails.

Proposition 3.4.3

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. Then

$$P\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} P(A_n) \leq \limsup_{n\to\infty} P(A_n) \leq P\left(\limsup_{n\to\infty} A_n\right)$$

Proof.

The middle inequality holds by definition, and the last inequality follows similarly to the first. We note that as $n \to \infty$, $\{\bigcap_{k=n}^{\infty} A_k\}_{n=1}^{\infty}$ are increasing to $\liminf_{n\to\infty} A_n$. Hence, by the monotone convergence theorem, we have

$$\begin{aligned}
P\left(\liminf_{n\to\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
&= \liminf_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
&\le \liminf_{n\to\infty} P(A_n)
\end{aligned}$$

where the final inequality follows by definition (if a limit exists, then it is equal to the liminf), and the final inequality follows from monotonicity. ■

Example 3.4.4

Consider the infinite fair coin tossing example in Example 3.4.2. Proposition 3.4.3 says that if $\{H_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ occurs infinitely often, then

$$P(H_n) \geq \frac{1}{2}$$

which is interesting but vague. To improve this result, we require a stronger result.

Theorem 3.4.5: The Borel-Cantelli Lemma

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. The following hold.

(i) If $\sum_{n=1}^{\infty} P(A_n)$ converges, then $P\left(\limsup_{n\to\infty} A_n\right) = 0$.

(ii) If $\sum_{n=1}^{\infty} P(A_n)$ diverges and $\{A_n\}_{n=1}^{\infty}$ are independent events, then $P\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof.

To see that (i) holds, assume that $\sum_{n=1}^{\infty} P(A_n)$ is convergent. Then for any $m \in \mathbb{N}$, we have by subadditivity,

$$P\left(\limsup_{n\to\infty} A_n\right) \le P\left(\bigcup_{k=m}^{\infty} A_k\right) \le \sum_{k=m}^{\infty} P(A_k)$$

which approaches to 0 as $m \to \infty$.

To see that (ii) holds, assume that $\sum_{n=1}^{\infty} P(A_n)$ diverges. Observe that

$$\left(\limsup_{n\to\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$$

so it suffices by countable subadditivity to show that $P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0$ for each $n \in \mathbb{N}$. Indeed, for $n, m \in \mathbb{N}$, we have by independence and the fact that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$,

$$\begin{aligned} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) &\leq P\left(\bigcap_{k=n}^{n+m} A_k^c\right) \\ &= \prod_{k=n}^{n+m} (1 - P(A_k)) \\ &\leq \prod_{k=n}^{n+m} e^{-P(A_k)} \\ &= e^{-\sum_{k=n}^{n+m} P(A_k)} \end{aligned}$$

which approaches 0 as $m \to \infty$. ■

This theorem proves its significance since it asserts that if $\{A_n\}_{n=1}^{\infty}$ are independent events, then $P\left(\limsup_{n\to\infty} A_n\right)$ is either 0 or 1, and it cannot be any other probability values.

Remark 3.4.6

Note that the independence assumption in Theorem 3.4.5 (ii) cannot be omitted. Indeed, consider the infinite fair coin tossing, and let $\{A_n\}_{n=1}^{\infty}$ such that for all $n \in \mathbb{N}$, $A_n = A_{n+1}$ be the events that the first coin comes up heads. Then $\{A_n\}_{n=1}^{\infty}$ are not independent and

$$P\left(\limsup_{n\to\infty} A_n\right) = \frac{1}{2}$$

Index