LECTURE NOTES
ENEE 620
RANDOM PROCESSES IN COMMUNICATION
AND CONTROL

Armand M. Makowski

1©2011 – 2022 by Armand M. Makowski
2Department of Electrical and Computer Engineering, University of Maryland, College
Park, MD 20742. E-mail: armand@umd.edu. Phone: (301) 405-6844
Chapter 0

Some notation, conventions
and terminology

In this preliminary chapter we briefly present the notation, terminology and convention to be used throughout this text.

0.1 Usual mathematical symbols

Throughout, we use $\mathbb{N}$ to denote the set $\{0, 1, \ldots\}$ of all non-negative integers, and write $\mathbb{N}_0$ to denote the set $\{1, 2, \ldots\}$ of all positive integers. We also write $\mathbb{R}$ to denote the set of all real numbers, while the notation $\mathbb{R}_+$ is reserved to represent the set $\{x \in \mathbb{R} : x \geq 0\}$ of all non-negative numbers. We introduce the extended real line to be the set $\mathbb{R} \cup \{-\infty, +\infty\}$, and we write $\mathbb{R}_+$ to denote the extended positive real line, namely $\mathbb{R}_+ = \mathbb{R} \cup \{+\infty\}$.

0.2 Countability vs. uncountability

A set $S$ is said to be countable if there is a one-to-one (or injective) mapping $f : S \to \mathbb{N}_0$ – In other words, the set $f(S)$ is a subset of $\mathbb{N}_0$. The countable set $S$ is said to be finite (resp. countably infinite) if $|f(S)| < \infty$ (resp. $|f(S)| = \infty$). We refer to a set that is not countable as being uncountable. When $|f(S)| < \infty$, say $|f(S)| = N$ for some positive integer $N$, the elements of $S$ can always be labelled so that $S = \{s_1, \ldots, s_N\}$. When $|f(S)| = \infty$, the elements of $S$ can always be labelled so that $S = \{s_1, \ldots, s_n, \ldots\}$ – Such labelings are not unique.
0.3 Displayed equations

0.4 Set theory

This section presents a brief review of some notions of Set Theory: We use \(\emptyset\) to denote the empty set. Throughout, with \(S\) an arbitrary non-empty set, let \(\mathcal{P}(S)\) denote the collection of all subsets of \(S\) (including the empty set) – We also refer to \(\mathcal{P}(S)\) as the power set of \(S\) (sometimes also denoted \(2^S\)).

**Elementary set-theoretic operations** With \(E\) and \(F\) subsets of \(S\), we write \(E \subseteq F\) when every element of \(E\) is also an element of \(F\), and refer to this situation by saying that \(E\) is contained in \(F\) or that \(E\) is a subset of \(F\) (resp. \(F\) is a superset of \(E\)).

The union and intersection of the subsets \(E\) and \(F\) are subsets of \(S\) which are denoted \(E \cup F\) and \(E \cap F\), respectively. They are defined by

\[
E \cup F \equiv \{s \in S : s \in E \text{ or } s \in F\}
\]

and

\[
E \cap F \equiv \{s \in S : s \in E \text{ and } s \in F\}
\]

We also define the following basic operations:

(i) the complement \(E^c\) of \(E\) (in \(S\)):

\[
E^c \equiv \{s \in S : s \notin E\}.
\]

(ii) the (set) difference \(E - F\):

\[
E - F \equiv E \cap F^c = \{s \in S : s \notin E\}.
\]

(iii) the symmetric difference \(E \Delta F\):

\[
E \Delta F \equiv (E - F) \cup (F - E) = (E \cap F^c) \cup (E^c \cap F).
\]

**De Morgan’s Laws** Let \(I\) denote an arbitrary index set. With \(\{E_i, \ i \in I\}\) a collection of subsets of \(S\), we have

\[
(\cup_{i \in I} E_i)^c = \cap_{i \in I} E_i^c
\]

and

\[
(\cap_{i \in I} E_i)^c = \cup_{i \in I} E_i^c.
\]
0.5. COLLECTIONS OF SETS

**Distributivity**  Let $I$ denote an arbitrary index set. With $\{E_i, i \in I\}$ a collection of subsets of $S$ and a subset $F$ of $S$, we have

$$\left( \bigcup_{i \in I} E_i \right) \cap F = \bigcup_{i \in I} (E_i \cap F)$$  \[\text{Set intersection is distributive over set union}\]

and

$$\left( \bigcap_{i \in I} E_i \right) \cup F = \bigcap_{i \in I} (E_i \cup F)$$  \[\text{Set union is distributive over set intersection}\]

0.5  Collections of sets

Since subsets of $S$ are elements of the power set $\mathcal{P}(S)$, we can think of a collection of subsets of $S$ as a subset of $\mathcal{P}(S)$. With this in mind we have the following

**Subsets**  If $\mathcal{H}_1$ and $\mathcal{H}_2$ are collections of subsets of $S$, we write $\mathcal{H}_1 \subseteq \mathcal{H}_2$ to express the fact that every subset of $S$ that belongs to $\mathcal{H}_1$ also belongs to $\mathcal{H}_2$. We then say that $\mathcal{H}_1$ is a *subset* of $\mathcal{H}_2$, or conversely that $\mathcal{H}_2$ is a *superset* of $\mathcal{H}_1$.

**Intersections and unions**  If $\{\mathcal{H}_i, i \in I\}$ is a non-empty family of collections of subsets of $S$, i.e., $\mathcal{H}_i \subseteq \mathcal{P}(S)$ for each $i$ in $I$, then their *intersection* $\bigcap_{i \in I} \mathcal{H}_i$ is the collection of subsets of $S$ given by

$$\bigcap_{i \in I} \mathcal{H}_i \equiv \{E \in \mathcal{P}(S) : E \in \mathcal{H}_i, i \in I\}.$$  

In other words, the collection $\bigcap_{i \in I} \mathcal{H}_i$ comprises all the subsets of $S$ that belong simultaneously to *each* of the collections $\{\mathcal{H}_i, i \in I\}$. In this definition the index set $I$ can be taken to be arbitrary.

0.6  Cartesian products

Let $S_a$ and $S_b$ be two arbitrary sets (possibly identical). The *Cartesian product* of $S_a$ and $S_b$, denoted $S_a \times S_b$, is the set of *ordered pairs* defined by

$$S_a \times S_b \equiv \{(s_a, s_b) : s_a \in S_a, s_b \in S_b\}.$$  

We refer to $S_a$ and $S_b$ as the *factors* of the Cartesian product $S_a \times S_b$. If $S_c$ is a third set (possibly identical to either $S_a$ or $S_b$), we *identify* $(S_a \times S_b) \times S_c$ with $S_a \times (S_b \times S_c)$ in the obvious way and write $S_a \times S_b \times S_c$ for either set. The generalization to more than two factors is obtained in an obvious manner.

In particular, it is customary to write the Cartesian product of $p$ copies of the same set $S_a$ as $S_a \times \ldots \times S_a$ or simply as $S_a^p$. 
Chapter 1

Modeling random experiments: The Kolmogorov model

A random experiment $\mathcal{E}$ is understood as an activity with the following characteristics: It typically has multiple possible outcomes, and the outcome of a realization of the experiment is revealed only after the experiment has been realized. Classical examples include the throw of a dice, the price of a commodity at the end of a trading day on some stock exchange, the temperature taken at noon on January 1 at the top of the Empire State Building, etc.

In these notes we use a widely accepted approach to modeling random experiments that is based on the measure-theoretic formalism proposed by Kolmogorov: According to this approach, a random experiment $\mathcal{E}$ is modeled through a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ where

- The set $\Omega$ lists all (elementary) outcomes (also known as samples) generated by the experiment $\mathcal{E}$; it is known as the sample space for the experiment.

- Events are collections of elementary outcomes, and so are subsets of $\Omega$. The collection of events to which likelihood of occurrence can be assigned is a collection $\mathcal{F}$ of events on $\Omega$. In many cases of interest one is forced for mathematical reasons to take $\mathcal{F}$ to be strictly smaller than the collection of all subsets of $\Omega$.

- The “likelihood” of occurrence of events is assigned only to the events in $\mathcal{F}$, and is given by means a probability measure $\mathbb{P}$ defined on $\mathcal{F}$.

These objects will be given precise mathematical meanings in what follows.
1.1 Fields and \( \sigma \)-fields

Throughout, with \( S \) a non-empty set, let \( \mathcal{S} \) denote a non-empty collection of subsets of \( S \), so that \( \mathcal{S} \subseteq \mathcal{P}(S) \).

**Definition 1.1.1**

The collection \( \mathcal{S} \) is said to be a *field* (also known as an *algebra* in some literature) on \( S \) if the conditions (F1)-(F3) hold where

(F1) \( \emptyset \in \mathcal{S} \).

(F2) Closed under complementarity: If \( E \in \mathcal{S} \), then \( E^c \in \mathcal{S} \).

(F3) Closed under union: If \( E \in \mathcal{S} \) and \( F \in \mathcal{S} \), then \( E \cup F \in \mathcal{S} \).

The De Morgan’s Laws have straightforward implications: The conditions (F1) and (F2) automatically imply that \( S \) is an element of the field \( \mathcal{S} \). Furthermore, (F2) and (F3) automatically imply

(F4) Closed under intersection: If \( E \in \mathcal{S} \) and \( F \in \mathcal{S} \), then \( E \cap F \in \mathcal{S} \)

(F5) Closed under differences: If \( E \in \mathcal{S} \) and \( F \in \mathcal{S} \), then \( E - F \in \mathcal{S} \), \( F - E \in \mathcal{S} \) and \( E \Delta F \in \mathcal{S} \).

Note that (F3) implies (is in fact equivalent to) the seemingly more general statement

(F3b) Closed under finite union: For each \( n = 1, 2, \ldots \), if \( E_1 \in \mathcal{S}, \ldots, E_n \in \mathcal{S} \), then \( \bigcup_{i=1}^{n} E_i \in \mathcal{S} \).

while (F4) implies (is in fact equivalent to) the seemingly more general statement

(F4b) Closed under finite intersection: For each \( n = 1, 2, \ldots \), if \( E_1 \in \mathcal{S}, \ldots, E_n \in \mathcal{S} \), then \( \bigcap_{i=1}^{n} E_i \in \mathcal{S} \).

For technical reasons that will soon become apparent a stronger notion is needed.

**Definition 1.1.2**

The non-empty collection of \( \mathcal{S} \) of subsets of \( S \) is a \( \sigma \)-field (also known as a \( \sigma \)-algebra) on \( S \) if

(F1) \( \emptyset \in \mathcal{S} \).
1.2. ADDITIVITY AND MEASURES

(F2) Closed under complementarity: If $E \in \mathcal{S}$, then $E^c \in \mathcal{S}$.

(F6) Closed under countable union: With $I$ a countable index set, if $E_i \in \mathcal{S}$ for each $i \in I$, then $\bigcup_{i \in I} E_i \in \mathcal{S}$.

Any $\sigma$-field is always a field since the additional property (F6) surmises (F3b) (which is itself equivalent to (F3)). Again, using De Morgan’s Laws we conclude under (F1) and (F2) that (F6) is equivalent to the following statement:

(F6b) Closed under countable intersection: With $I$ a countable index set, if $E_i \in \mathcal{S}$ for each $i \in I$, then $\bigcap_{i \in I} E_i \in \mathcal{S}$.

Any set $S$ always carries at least two $\sigma$-fields, namely the trivial $\sigma$-field $\{\emptyset, S\}$ and the full $\sigma$-field $\mathcal{P}(S)$. With an arbitrary set $S$ and a $\sigma$-field $\mathcal{S}$ on $S$, it is customary to refer to the pair $(S, \mathcal{S})$ as a measurable space. This is meant to suggest that it is now possible to “measure” the sets in $\mathcal{S}$ by means of a measure defined on $S$, an idea formalized in the next section.

1.2 Additivity and measures

Let $\mathcal{S}$ denote a non-empty collection of subsets of some non-empty set $S$. Measuring the sets in $\mathcal{S}$ means that a notion of “size” (also referred to as “length” or “volume” or “weight” depending on the context) can be associated with every such set. This is done through a set function which maps any set $S$ in $\mathcal{S}$ to a non-negative (possibly infinite) value $\mu(S)$. Of course we expect such a set function to satisfy some natural properties. Additivity is the most obvious one as it reflects the natural idea that the size of an object can be evaluated as the sum of the sizes of its “non-overlapping” components; this is formalized next.

Definition 1.2.1 With arbitrary index set $I$, the subsets $\{E_i, i \in I\}$ of $S$ are said to be pairwise disjoint, or simply disjoint, if

$$E_i \cap E_j = \emptyset, \quad i \neq j, \quad i, j \in I.$$
CHAPTER 1. MODELING RANDOM EXPERIMENTS: THE KOLMOGOROV MODEL

Definition 1.2.2

Given a collection \( S \) of subsets of \( S \), a set function \( \mu : S \to [0, \infty] \) is finitely additive, or simply additive, on \( S \) if for any finite collection \( \{ E_i, i \in I \} \) of elements in \( S \) we have

\[
\mu \left( \bigcup_{i \in I} E_i \right) = \sum_{i \in I} \mu \left( E_i \right)
\]

whenever the sets \( \{ E_i, i \in I \} \) are disjoint, and their union \( \bigcup_{i \in I} E_i \) belongs to \( S \).

The natural setting for this definition is for \( S \) to be a field on \( S \) since then the union set \( \bigcup_{i \in I} E_i \) automatically belongs to \( S \) when the sets in the finite collection \( \{ E_i, i \in I \} \) are elements of the field \( S \).

In order to deal with situations where the sample space is countably infinite or uncountable, we extend the definition of a finitely additive set function in very much the same way that we extended the notion of a field to that of a \( \sigma \)-field – This is done by allowing the additive evaluation of unions of countably many, not just finitely many, disjoint events.

Definition 1.2.3

Given a collection \( S \) of subsets of \( S \), a set function \( \mu : S \to [0, \infty] \) is countably additive, or simply \( \sigma \)-additive, if for any countable collection \( \{ E_i, i \in I \} \) of elements in \( S \) we have

\[
\mu \left( \bigcup_{i \in I} E_i \right) = \sum_{i \in I} \mu \left( E_i \right)
\]

whenever the sets \( \{ E_i, i \in I \} \) are disjoint, and their union \( \bigcup_{i \in I} E_i \) belongs to \( S \).

This time the natural setting for this definition is for \( S \) to be a \( \sigma \)-field on \( S \) since then the set \( \bigcup_{i \in I} E_i \) automatically belongs to \( S \) when the countably many sets \( \{ E_i, i \in I \} \) are elements of the \( \sigma \)-field \( S \). On the other hand, according to Definition 1.2.3 a countably additive set function \( \mu : S \to [0, \infty] \) when defined on a field \( S \) is automatically finitely additive there.

Definition 1.2.4

Let \( S \) be an arbitrary non-empty set equipped with a \( \sigma \)-field \( S \). A \( \sigma \)-additive measure \( \mu \) on \( S \) is a set function \( \mu : S \to [0, \infty] \) which satisfies the properties (M1)-(M2) where

(M1) \( \mu \left( \emptyset \right) = 0 \).
1.3. PROBABILITY MEASURES

(M2) $\sigma$-additivity: For any countable collection $\{E_i, i \in I\}$ of disjoint subsets in $\mathcal{S}$, we have

$$\mu \left( \bigcup_{i \in I} E_i \right) = \sum_{i \in I} \mu \left[ E_i \right].$$

A $\sigma$-additive measure is often referred simply as a measure; this terminology always assumes that its domain of definition $\mathcal{S}$ is a $\sigma$-field. The qualifier “on $\mathcal{S}$” is usually dropped once it is clear from the context what is the $\sigma$-field $\mathcal{S}$ on $\mathcal{S}$ being used throughout the discussion. However, sometimes the qualifier “on $(\mathcal{S}, \mathcal{S})$” is added when there might be ambiguity as to the measurable space being considered.

A measure is said to be finite if $\mu [S] < \infty$, in which case (M1) is automatically satisfied as a consequence of (M2) since $\mu [S] = \mu [S] + \mu [\emptyset]$ by additivity on account of the obvious relations $S = S \cup \emptyset$ and $S \cap \emptyset = \emptyset$.

There are important situations where $\mu [S] = \infty$ but much of measure theory can still be developed as in the finite case through localization arguments. This arises when the measure $\mu : \mathcal{S} \to [0, \infty]$ is $\sigma$-finite in the following sense: There exists a sequence of sets $\{E_n, n = 1, 2, \ldots\}$ in the $\sigma$-field $\mathcal{S}$, said sequence being monotone increasing, i.e., $E_n \subset E_{n+1}$ for all $n = 1, 2, \ldots$, which “exhausts” $\mathcal{S}$ in that $\bigcup_{n=1}^{\infty} E_n = \mathcal{S}$ and for which $\mu [E_n] < \infty$ for all $n = 1, 2, \ldots$.

1.3 Probability measures

Specializing Definition 1.2.4 we obtain the notion of a probability measure, a notion that will occupy a central place in further developments.

**Definition 1.3.1**

Let $\mathcal{S}$ be an arbitrary non-empty set equipped with a $\sigma$-field $\mathcal{S}$. A probability measure $\mu : \mathcal{S} \to \mathbb{R}_+$ on $\mathcal{S}$ is a finite measure on $\mathcal{S}$ with $\mu [S] = 1$.

Collecting earlier definitions and remarks we readily see that the set function $\mu : \mathcal{S} \to \mathbb{R}_+$ is a probability measure on $(\mathcal{S}, \mathcal{S})$ if and only if it satisfies the following properties:

(P1) $\mu [S] = 1$.

(P2) $\sigma$-additivity: For any countable collection $\{E_i, i \in I\}$ of disjoint subsets in $\mathcal{S}$, we have

$$\mu \left( \bigcup_{i \in I} E_i \right) = \sum_{i \in I} \mu \left[ E_i \right].$$
As mentioned earlier, the condition $\mu[S] = 1$ implies $\mu[\emptyset] = 0$. Moreover, if $E$ is any event in the $\sigma$-field $S$, then its complement $E^c$ is also in the $\sigma$-field $S$ with $E \cup E^c = S$, whence

$$
\mu[E] + \mu[E^c] = \mu[S] = 1
$$

by additivity. It follows that

$$
\mu[E^c] = 1 - \mu[E], \quad E \in S
$$

and

$$
0 \leq \mu[E] \leq 1, \quad E \in S.
$$

In other words, \{\mu[E], E \in S\} \subseteq [0, 1] and a probability measure is a set function $\mu : S \to [0, 1]$, note merely $\mu : S \to \mathbb{R}_+$.

### 1.4 Probability models

As likelihood assignments are implemented through probability measures, we are now ready to introduce the basic model that we will adopt in the study of random phenomena (with the usual change of notation).

**Definition 1.4.1**

A *probability model* for the random experiment $\mathcal{E}$ is a triple $(\Omega, \mathcal{F}, P)$ where the set $\Omega$ is the sample space for the experiment, $\mathcal{F}$ is a $\sigma$-field of events on $\Omega$ and $P$ is a probability measure on $(\Omega, \mathcal{F})$ (or simply on $\mathcal{F}$).

We refer to $(\Omega, \mathcal{F}, P)$ as a *probability space* (or as a *probability triple*). An event $E$ in $\mathcal{F}$ such that $P[E] = 1$ is called a *certain* event, whereas an event $E$ in $\mathcal{F}$ such that $P[E] = 0$ is called a *null* event. Next we present simple, yet useful, consequences of the definitions (F1)-(F5) and (P1)-(P2); proofs are elementary and left to the interested reader as exercises [Exercise 1.6] – Some have already been given.

Given a probability triple $(\Omega, \mathcal{F}, P)$, with events $E$ and $F$ in $\mathcal{F}$, we have

(i) Complementarity:

$$
P[E^c] = 1 - P[E].
$$

(ii) Generalizing additivity:

$$
P[E \cup F] = P[E] + P[F] - P[E \cap F]
$$

so that

$$
P[E \cup F] \leq P[E] + P[F].
$$
1.5 Discrete Probability Models and PMFs

(iii) Monotonicity (I):
\[ P[F] = P[F - E] + P[E], \quad E \subseteq F. \]

(iv) Monotonicity (II):
\[ P[E] \leq P[F], \quad E \subseteq F. \]

(v) Monotonicity (III):
\[ 0 \leq P[E] \leq 1. \]

1.5 Discrete probability models and pmfs

In many applications a major question is concerned with determining the probability measure \( P \) that captures the salient features of the experiment \( E \) under consideration once its sample space \( \Omega \) has been identified. This requires that the \( \sigma \)-field \( F \) of events be judiciously chosen.

A situation of particular importance arises when \( \Omega \) is countable, in which case it is customary to take \( F = \mathcal{P}(\Omega) \) – This choice reflects the natural desire to assign the likelihood of occurrence to the individual outcomes \( \{\omega\}, \omega \in \Omega \) (so that anticipating on the material of Section 1.7 we must have \( \sigma(\{\omega\}, \omega \in \Omega) = \mathcal{P}(\Omega) \) [Exercise 1.14]). We refer to such models as discrete probability models.

As we now argue, specifying \( P \) on \((\Omega, \mathcal{P}(\Omega))\) is equivalent to specifying

\[
\{P[\{\omega\}], \omega \in \Omega\}.
\]

(1.1)

Indeed, if \( P \) has been specified on \((\Omega, \mathcal{P}(\Omega))\), then obviously the values (1.1) are known since \( \{\omega\} \) is (an event) in \( \mathcal{P}(\Omega) \) for every sample \( \omega \) in \( \Omega \). Conversely, if the values \( \{P[\{\omega\}], \omega \in \Omega\} \) were only available, then the obvious relation

\[ E = \bigcup_{\omega \in E} \{\omega\}, \quad E \in \mathcal{P}(\Omega) \]

implies

\[ P[E] = \sum_{\omega \in E} P[\{\omega\}], \quad E \in \mathcal{P}(\Omega) \]

by the \( \sigma \)-additivity of \( P \) since every subset of the countable sample space \( \Omega \) is necessarily countable. This shows that the values \( \{P[\{\omega\}], \omega \in \Omega\} \) indeed uniquely specify \( P \) on the whole \( \sigma \)-field \( \mathcal{P}(\Omega) \), an observation which leads to the following elementary fact.
Fact 1.5.1 With $\Omega$ a countable set, any probability measure $P$ on the $\sigma$-field $P(\Omega)$ can be uniquely represented by a collection $\{p(\omega), \omega \in \Omega\}$ satisfying

\begin{equation}
0 \leq p(\omega) \leq 1, \quad \omega \in \Omega \quad \text{and} \quad \sum_{\omega \in \Omega} p(\omega) = 1,
\end{equation}

with the identification $P[\{\omega\}] = p(\omega)$ for each $\omega$ in $\Omega$. We necessarily have

\begin{equation}
P[E] = \sum_{\omega \in E} p(\omega), \quad E \in P(\Omega).
\end{equation}

We often refer to a collection $\{p(\omega), \omega \in \Omega\}$ satisfying (1.2) as a probability mass function (pmf) on $\Omega$, written $(p(\omega), \omega \in \Omega)$.

1.6 Uniform probability assignments

Let $\Omega$ be an arbitrary set to be used as the sample space of a probabilistic experiment $E$ where outcomes are known or believed to be equally likely to occur – In many literatures an outcome of $\Omega$ so selected is said to be selected at random. We will avoid this usage and use instead the more accurate terminology whereby the outcomes are uniformly generated. A natural question is how to construct the corresponding probability measure $P$, hereafter referred to as the uniform probability measure. Obviously such a construction also requires that we simultaneously identify the appropriate $\sigma$-field $F$ of events on which $P$ is defined.

In a first step it may seem intuitively reasonable to start building this uniform probability measure by assigning the same probability of occurrence to all outcomes. This would necessarily require that $\{\{\omega\}, \omega \in \Omega\} \subseteq F$, i.e.,

\begin{equation}
\{\omega\} \in F, \quad \omega \in \Omega
\end{equation}

with

\begin{equation}
P[\{\omega\}] = p, \quad \omega \in \Omega
\end{equation}

for some $p$ in $[0, 1]$. Again, anticipating on the material of Section 1.7 we must have $\sigma(\{\{\omega\}, \omega \in \Omega\}) \subseteq F$.

Under the requirement (1.4) any countable subset $E$ of $\Omega$ must belong to the $\sigma$-field $F$ as a consequence of the decomposition $E = \bigcup_{\omega \in E} \{\omega\}$. Thus, by $\sigma$-additivity we conclude that

\begin{equation}
P[E] = \sum_{\omega \in E} P[\{\omega\}], \quad E \subseteq \Omega \quad \text{Countable.}
\end{equation}

Several cases arise:
1.6. UNIFORM PROBABILITY ASSIGNMENTS

Finite case \((\Omega < \infty)\) The set \(\Omega\) contains a finite number of elements, say \(\Omega = \{\omega_1, \ldots, \omega_N\}\) for some finite \(N\). As pointed out in Section 1.5 the requirement \((1.4)\) leads to \(\mathcal{F} = \mathcal{P}(\Omega)\). Using \((1.5)\) into \((1.6)\) we get

\[
P[E] = \sum_{\omega \in E} P[\{\omega\}] = |E|p, \quad E \in \mathcal{P}(\Omega)
\]

whence \(p = |\Omega|^{-1}\) upon taking \(E = \Omega\) in this last relation. It immediately follows that

\[
P[E] = \frac{|E|}{|\Omega|}, \quad E \in \mathcal{P}(\Omega).
\]

Much of elementary Probability Theory is concerned with computing such probabilities through combinatorial arguments that help evaluate the size of various subsets (e.g., \(E\)) of a discrete set (e.g., \(\Omega\)).

Countably infinite case \((|\Omega| = \infty)\) The set \(\Omega\) contains countably infinite many elements, say \(\Omega = \{\omega_n, \ n = 1, 2, \ldots\}\) for some labeling \(\mathbb{N}_0 \to \Omega: n \to \omega_n\). Again, as pointed out in Section 1.5 the requirement \((1.4)\) leads to \(\mathcal{F} = \mathcal{P}(\Omega)\). The same argument as above, based on \((1.5)\) and \((1.6)\), shows that

\[
P[E] = |E|p \leq 1, \quad E \in \mathcal{P}(\Omega)
\]

Now it is always possible to select a sequence \(\{E_n, \ n = 1, 2, \ldots\}\) of subsets of \(\Omega\) such that \(|E_n| = n\) for all \(n = 1, 2, \ldots\) – Indeed, with the labeling introduced earlier, just take \(E_n = \{\omega_1, \ldots, \omega_n\}\) in which case \(|E_n| = n\). Applying \((1.8)\) with \(E = E_n\) we conclude \(p \leq n^{-1}\) for all \(n = 1, 2, \ldots\), whence \(p = 0\). A contradiction immediately arises: Indeed, by virtue of \((1.6)\) (with \(E = \Omega\)), we get \(P[\Omega] = \sum_{\omega \in \Omega} p = 0\), and yet we must have \(P[\Omega] = 1\) because \(P\) is a probability measure. In other words, on a discrete sample set \(\Omega\) with \(|\Omega| = \infty\) it is not possible to construct a probability measure that satisfies the uniformity constraint \((1.5)\).

Uncountably infinite case When \(\Omega\) is uncountable, the same arguments as above will still show that \(p = 0\), and the conclusion

\[
P[E] = 0, \quad E \subseteq \Omega
\]

again follows from \((1.6)\) by \(\sigma\)-additivity.
CHAPTER 1. MODELING RANDOM EXPERIMENTS: THE KOLMOGOROV MODEL

What can we say concerning \( P[E] \) if the subset \( E \) is not countable? While the decomposition \( E = \bigcup_{\omega \in E} \{\omega\} \) always holds for any subset \( E \) of \( \Omega \), there is no guarantee that

\[
P[E] = \sum_{\omega \in E} P[\{\omega\}], \quad E \subseteq \Omega \text{ Uncountable}
\]

since \( P \) is only required to be \( \sigma \)-additive. In fact, were this last relationship indeed hold for every subset of \( \Omega \) (including \( \Omega \)) we would have to conclude that \( P[E] = 0 \) for every subset of \( \Omega \) (including \( \Omega \), hence already a contradiction) This would certainly define a measure on \( \mathcal{P}(\Omega) \), namely the zero measure, definitely not a probability measure on \( \mathcal{P}(\Omega) \).

This discussion strongly suggests that when defining probability measures on non-countable sample spaces \( \Omega \), under the uniformity constraint (1.4) it may not be feasible to take \( \mathcal{F} = \mathcal{P}(\Omega) \) – This will be further discussed in Chapter ?? This can be traced back to the fact that the \( \sigma \)-additivity of \( P \) on \( \mathcal{P}(\Omega) \) imposes too many constraints – They cannot all be simultaneously satisfied if \( P \) is to be defined on \( \mathcal{P}(\Omega) \), thereby forcing a reduction of \( \mathcal{P}(\Omega) \) to a strictly smaller \( \sigma \)-field!

However, the analysis so far does not preclude the possibility of constructing a probability measure \( P \) which reflects constraints naturally associated with uniform selection other than (1.4) and (1.5) – This is illustrated on two examples, namely infinitely many coin tosses of a fair coin in Section ?? and selecting a point at random in the interval \([0, 1]\) in Section ??.. However, such constructions will have to be done on a \( \sigma \)-field strictly smaller than \( \mathcal{P}(\Omega) \).

The reader may wonder as to why the conclusions reached in the countably infinite and uncountable cases differ: In the countable case the equivalence embedded in Fact 1.5.1 forces the construction of the desired uniform probability measure to pass through the constraints (1.4)-(1.5). While this leads to a complete characterization, namely (1.7), when \( \Omega \) contains a finite number of samples, the situation is quite different in the countably infinite case: There the constraints (1.4)-(1.5) are now incompatible with \( \Omega \) being countably infinite. In the non-countable case, the construction of the desired uniform probability measure cannot pass through the constraints (1.4)-(1.5) (which are too weak as will be illustrated in Section ??), leaving open the possibility that additional constraints reflecting uniformity could possibly be added to characterize the desired probability measure.
1.7 Generating σ-fields

On a number of occasions it will be helpful to consider the smallest σ-field containing a given collection of subsets of a non-empty set \( S \). The following elementary fact provides the basis for this notion; its proof is left as an exercise [Exercise 1.11].

**Fact 1.7.1** If \( \{ F_i, i \in I \} \) is a non-empty family of σ-fields on \( S \) (with \( I \) arbitrary, countable or not), then the intersection \( \bigcap_{i \in I} S_i \) is a σ-field on the set \( S \).

Using Fact 1.7.1 it is a simple matter to define the desired concept by leveraging the following easy result [Exercise 1.12].

**Lemma 1.7.1** Let \( G \) denote a non-empty collection of subsets of \( S \). There exists a unique σ-field on \( S \), denoted \( \sigma(G) \), with the following properties:

(i) The σ-field \( \sigma(G) \) contains \( G \);

(ii) The σ-field \( \sigma(G) \) is minimal in the sense that any other σ-field \( S \) on \( S \) containing \( G \) must necessarily contain \( \sigma(G) \), i.e., \( \sigma(G) \subseteq S \).

We refer to the σ-field \( \sigma(G) \), whose existence is established in Lemma 1.7.1, as the σ-field on \( S \) generated by \( G \). In fact it is easy to see that \( \sigma(G) \) coincides with the σ-field \( \bigcap_{i \in I} G_i \) where \( \{ G_i, i \in I \} \) denotes the family of all the σ-fields on \( S \) containing the collection \( G \) – Note that \( \{ G_i, i \in I \} \) is never empty as it contains the power set \( \mathcal{P}(S) \).

**Definition 1.7.1**

Let \( G \) and \( S \) be two collections of subsets of \( S \) with \( G \subseteq S \). If \( S \) is a σ-field on \( S \) with \( S = \sigma(G) \), we say that \( G \) generates the σ-field \( S \), or equivalently, that \( G \) is a generating family (or a generator) for \( S \).

---

The following fact is elementary [Exercise 1.13].

**Fact 1.7.2** If \( G_1 \) and \( G_2 \) are two collections of subsets of \( S \) such that \( G_1 \subseteq G_2 \), then \( \sigma(G_1) \subseteq \sigma(G_2) \). Moreover, if \( G_2 \) is already a σ-field, then \( \sigma(G_2) = G_2 \) and \( \sigma(G_1) \subseteq G_2 \).

1.8 Exercises

**Ex. 1.1** Let \( S \) be a σ-field on some non-empty set \( S \) with a finite number of elements, i.e., \( |S| < \infty \).
a. Let $S^*$ denote the collection of all non-empty elements of $S$ which do not contain another non-empty element of $S$. Explain how $S$ can be generated from $S^*$ – We can think of $S^*$ as the “atoms” of $S$ [HINT: Remember that $S$ is a $\sigma$-field on $S$].

b. Using Part a show that we necessarily have $|S| = 2^m$ with $|S^*| = m$.

c. Claim: Any $\sigma$-field on a non-empty finite set $S$ necessarily has $2^m$ subsets of $S$ in it for some positive integer $m$.

Ex. 1.2 Let $\mathcal{H}$ be a field on some set $S$. For any additive set function $\mu : \mathcal{H} \to [0, \infty]$ show that $\mu([0]) = 0$ as soon as there exists $H$ in $\mathcal{H}$ such that $\mu(H) < \infty$.

Ex. 1.3 In Definition 1.2.2 show that it suffices to check that the simpler pairwise conditions
\[\mu(E \cup F) = \mu(E) + \mu(F), \quad E, F \in S \quad E \cap F = \emptyset\]
hold.

Ex. 1.4 Let $S$ denote a countable set. With $\mathcal{F} = \mathcal{P}(S)$, define the set function $\mu : \mathcal{F} \to \mathbb{R}_+$ by
\[\mu(E) = |E|, \quad E \in \mathcal{F}\]
where $|E|$ denotes the number of elements in $E$. Show that the set function $\mu : \mathcal{F} \to \mathbb{R}_+$ is a measure on $\mathcal{F}$ – It is known as the counting measure.

Ex. 1.5 Let $S$ be a countably infinite set, say $S = \mathbb{N}$. Define the collection $\mathcal{F}$ of subsets of $S$ to be $\mathcal{F} \equiv \{F \subseteq S : \text{Either } |F| < \infty \text{ or } |F^c| < \infty \}$.

a. Show that $\mathcal{F}$ is an algebra on $S$. Is it a $\sigma$-algebra on $S$? Explain.

b. Define the mapping $\mu : \mathcal{F} \to \mathbb{R}_+$ by
\[\mu(E) \equiv \begin{cases} 0 & \text{if } |E| < \infty \\ 1 & \text{if } |E^c| < \infty. \end{cases}\]
Show that $\mu$ is finitely additive. Is $\mu$ also $\sigma$-additive on $\mathcal{F}$? – Prove or give a counterexample!

Ex. 1.6 Give proofs to the elementary properties (i)-(v) of probability models given in Section 1.4.

Ex. 1.7 Additional elementary properties of a probability measure: Given a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, with events $E$, $F$ and $G$, we have
\[\mathbb{P}(E \cap F) \geq \mathbb{P}(E) + \mathbb{P}(F) - 1,\]
and 
\[ \mathbb{P} [E \Delta F] = \mathbb{P} [E] + \mathbb{P} [F] - 2\mathbb{P} [E \cap F] \]

where \( E \Delta F \) denotes the symmetric difference of \( E \) and \( F \) (defined as \( E \Delta F = (E \cap F^c) \cup (F \cap E^c) \)). Furthermore, the following “triangle inequality”

\[ \mathbb{P} [E \Delta G] \leq \mathbb{P} [E \Delta F] + \mathbb{P} [F \Delta G] \]

holds.

**Ex. 1.8** Let \( S \) be an uncountable set, say \( S = \mathbb{R} \). Define the collection \( \mathcal{F} \) of subsets of \( S \) to be \( \mathcal{F} \equiv \{ E \subseteq S : E \text{ is countable or } F^c \text{ is countable} \} \) where countable means here either finite or countably infinite.

a. Show that \( \mathcal{F} \) is an algebra on \( S \).

b. Is \( \mathcal{F} \) a \( \sigma \)-algebra on \( S \)? Prove or give a counterexample!

c. Define the set function \( \mu : \mathcal{F} \to \mathbb{R}_+ \) by

\[ \mu [E] \equiv \begin{cases} 
0 & \text{if } E \text{ is countable} \\
1 & \text{if } E^c \text{ is countable.}
\end{cases} \]

Is this set function \( \mu : \mathcal{F} \to \mathbb{R}_+ \) a probability measure on \( \mathcal{F} \)? Prove or give a counterexample!

**Ex. 1.9** Let \( S \) denote a finite set, say \( S = \{1, \ldots, n\} \) for some positive integer \( n \). The random experiment \( \mathcal{E} \) consists in selecting uniformly an ordered pair \( A \) and \( B \) of (possibly empty) subsets of \( S \).

a. Construct an probability model for this experiment – Clearly specify the sample space \( \Omega \), the \( \sigma \)-field \( \mathcal{F} \) of events and the probability assignment \( \mathbb{P} \). Using the ideas developed in Section 1.6, compute the following probabilities:

b. For any subset \( B \) of \( S \), compute the probability of the event \( \mathcal{I}_B \) given by

\[ \mathcal{I}_B \equiv \{ (A, B) : (A, B) \in \mathcal{P}(S) \times \mathcal{P}(S) : A \subseteq B \} \]

c. Compute the probability of the event \( \mathcal{I} \) given by

\[ \mathcal{I} \equiv \{ (A, B) : (A, B) \in \mathcal{P}(S) \times \mathcal{P}(S) : A \subseteq B \} \]

[HINT: Note that \( \mathcal{I} = \bigcup_{B \subseteq S} \mathcal{I}_B \)]. See Exercise 2.27 for another take on this problem using conditional probabilities.
Ex. 1.10 Completing a probability space: Given a probability triple $(\Omega, \mathcal{F}, P)$, let $\mathcal{N}$ denote the collection of all null events (under $P$), i.e., $\mathcal{N} \equiv \{ N \in \mathcal{F} : P[N] = 0 \}$. Consider now the collection $\mathcal{N}^*$ of all subsets of $\Omega$ that are subsets of $P$-null events, i.e.,

$$\mathcal{N}^* \equiv \{ M \in \mathcal{P}(\Omega) : M \subseteq N \text{ for some } N \in \mathcal{N} \} = \cup_{N \in \mathcal{N}} \mathcal{P}(N).$$

Subsets in $\mathcal{N}^*$ are not necessarily events in $\mathcal{F}$.

a. Show that the collection $\mathcal{F}^* \equiv \{ E \cup M : E \in \mathcal{F}, M \in \mathcal{N}^* \}$ is also a $\sigma$-field on $\Omega$ (which contains $\mathcal{F}$).

b. Define the set function $P^* : \mathcal{F}^* \to [0, 1]$ by

$$P^*[E^*] \equiv P[E], \quad E^* = E \cup M, \quad E \in \mathcal{F}, M \in \mathcal{N}^*.$$ 

Show that this definition is well posed in the following sense: If $E^*$ admits the two representations $E_1 \cup M_1$ and $E_2 \cup M_2$ with $E_k \in \mathcal{F}$ and $M_k \in \mathcal{N}^*$, $k = 1, 2$, then $P[E_1] = P[E_2]$, thereby yielding an unambiguous value for $P^*[E^*]$ [HINT: Make use of the following observation: If for each $k = 1, 2$, $M_k \subseteq N_k$ where $N_k$ is an element of $\mathcal{N}$, then the equality $E_1 \cup M_1 = E_2 \cup M_2$ implies the inclusions $E_1 \subseteq M_2 \cup N_2$ and $E_2 \subseteq M_1 \cup N_1$].

c. Show that the set function $P^* : \mathcal{F}^* \to [0, 1]$ (which is well defined as per Part b) is a probability measure on $\mathcal{F}^*$ which coincides with $P$ on $\mathcal{F}$.

d. Show that the probability measure $P^*$ is complete on $\mathcal{F}^*$ in the sense that if $P^*[E^*] = 0$ for some $E^* \in \mathcal{F}^*$, then for any subset $E^{**}$ of $E^*$, it holds that $E^{**}$ belongs to $\mathcal{F}^*$ with $P^*[E^{**}] = 0$.

Ex. 1.11 Given a non-empty family $\{ S_i, i \in I \}$ of $\sigma$-fields (resp. fields) on some arbitrary set $S$, show that the collection $\cap_{i \in I} S_i$ is a $\sigma$-field (resp. field) on $S$.

Ex. 1.12 Prove Lemma 1.7.1.

Ex. 1.13 Prove Fact 1.7.2.

Ex. 1.14 Let $S$ be an arbitrary non-empty set. Let $\mathcal{G}$ denote the collection of all its singletons, namely $\mathcal{G} = \{ \{ s \} : s \in S \}$.

a. If $S$ is a finite set, what is $\sigma(\mathcal{G})$?

b. If $S$ is a countably infinite set, what is $\sigma(\mathcal{G})$?

c. If $S$ is an uncountably infinite set, what is $\sigma(\mathcal{G})$?
Ex. 1.15 With $S$ an arbitrary non-empty set, let $\mathcal{G}$ be a collection of subsets of $S$. If $E$ denotes a subset of $S$, define the *trace of $\mathcal{G}$ on $E$* as the collection $\mathcal{G}_E$ of subsets of $E$ given by

$$\mathcal{G}_E \equiv \{ G \cap E : G \in \mathcal{G} \}.$$ 

a. Show that $\mathcal{G}_E$ is a field (resp. a $\sigma$-field) on $E$ whenever $\mathcal{G}$ is a field (resp. $\sigma$-field) on $E$ regardless of whether $E$ is an element of $\mathcal{G}$.

b. Show that generating the smallest $\sigma$-field and taking a trace are commutative operations, i.e., $\sigma(\mathcal{G}_E) = (\sigma(\mathcal{G}))_E$. 

Chapter 2

Elementary Probability Theory

In Chapter 1 we introduced the notion of a probability model for a random experiment in the sense of Kolmogorov as a triple \((\Omega, \mathcal{F}, P)\) where \(\Omega\) lists all elementary outcomes of the experiment, and the collection \(\mathcal{F}\) identifies the subsets of \(\Omega\) (or events) whose likelihood can be evaluated by means of the probability measure \(P\) defined on \(\mathcal{F}\).

In the present chapter we present some of the most basic concepts often found in textbooks covering elementary Probability Theory. Throughout we are given a probability triple \((\Omega, \mathcal{F}, P)\) which is held fixed during the discussion.

2.1 Bounding probabilities

With \(I\) a finite index set, let \(\{E_i, i \in I\}\) denote any collection of events in \(\mathcal{F}\). At this point the reader may wonder how to evaluate the probability of the union \(\bigcup_{i \in I} E_i\) when the events are not disjoint (since in that case it is unclear how to invoke \(\sigma\)-additivity).

A formula by Poincaré The next result is attributed to Poincaré and gives a formal answer to this question. It states that

\[
\mathbb{P}\left[\bigcup_{i \in I} E_i\right] = \sum_{k=1}^{\vert I \vert} (-1)^{k-1} \left( \sum_{j_1 < j_2 < \ldots < j_k} \mathbb{P}\left[\bigcap_{\ell=1}^{k} E_{j_\ell}\right] \right)
\]

(2.1)

where for each \(k = 1, \ldots, \vert I \vert\), the summation \(\sum_{j_1 < j_2 < \ldots < j_k}\) is over all ordered \(k\)-uples drawn from \(I\). This formula is an expression of the Inclusion-Exclusion Principle.
By complementarity this expression is often used in the form
\[
P[\cap_{i \in I} E_i^c] = 1 - P[\cup_{i \in I} E_i]
\]
with the understanding that the term corresponding to \( k = 0 \) is set to 1. The expressions (2.1) and (2.2) are easily derived by induction on the size of the cardinality \( |I| \) [Exercise 2.1].

While exact, the expressions (2.1) and (2.2) are often too unwieldy to be useful. Instead only the upper and lower bounds given next suffice in many cases; they are also established by induction on the size \( |I| \) [Exercise 2.3].

**Boole’s inequality** This bound, also known as the *union bound*, is commonly used in Information Theory and theoretical Computer Science, and states that
\[
P[\cup_{i \in I} E_i] \leq \sum_{i \in I} P[E_i].
\]
The union bound (2.3) also holds when \( I \) is countably infinite [Exercise 4.1].

**The Bonferroni Inequality** This bound gives a *lower* bound on the probability \( P[\cup_{i \in I} E_i] \) in the form
\[
\sum_{i \in I} P[E_i] - \sum_{i,j \in I: i < j} P[E_i \cap E_j] \leq P[\cup_{i \in I} E_i].
\]
Combining the inequalities (2.3) and (2.4) we get
\[
\sum_{i \in I} P[E_i] - \sum_{i,j \in I: i < j} P[E_i \cap E_j] \leq P[\cup_{i \in I} E_i] \leq \sum_{i \in I} P[E_i].
\]
This opens the door to the possibility that \( P[\cup_{i \in I} E_i] \) might be well approximated by \( \sum_{i \in I} P[E_i] \) whenever the term \( \sum_{i,j \in I: i < j} P[E_i \cap E_j] \) is smaller than \( \sum_{i \in I} P[E_i] \) in a suitable sense. This idea is commonly used in many settings. Sometimes it is more convenient to express these two inequalities in the following equivalent form
\[
0 \leq \sum_{i \in I} P[E_i] - P[\cup_{i \in I} E_i] \leq \sum_{i,j \in I: i < j} P[E_i \cap E_j].
\]

**2.2 Independence**

The notion of independence introduced next is key to probabilistic modeling. It is perhaps what makes Probability Theory not just a special case but rather a very rich subarea of Measure Theory.
2.2. INDEPENDENCE

In this section we consider a collection \( \{E_i, i \in I\} \) of events in \( \mathcal{F} \) where \( I \) is an arbitrary index set, and present the several notions of independence commonly discussed.

**Pairwise independence** The events \( \{E_i, i \in I\} \) are said to be pairwise independent if the conditions

\[
P[E_i \cap E_j] = P[E_i]P[E_j], \quad i \neq j, \quad i, j \in I
\]

all hold. With \( I \) finite, this constitutes a set of \( \frac{|I|(|I|-1)}{2} \) conditions. When considering only two events, i.e., \( |I| = 2 \), this set of conditions reduces to a single condition, in which case the qualifier “pairwise” is dropped and the two events are simply said to be independent.

The terminology may be misleading. Indeed, *if two events are independent, it does not necessarily mean that their outcomes are not influencing each other in any way*; see Exercise 2.8 for an illustration of this point.

**Mutual independence (with \( I \) finite)** The events \( \{E_i, i \in I\} \) are said to be mutually independent if the conditions

\[
P[\cap_{j \in J} E_j] = \prod_{j \in J} P[E_j], \quad J \subset I, |J| \geq 1
\]

are all satisfied – This represents \( 2^{|I|} - (|I|+1) \) non-trivial conditions. In Exercises 2.6 and 2.7 situations are given (with \( |I| = 3 \) so that \( 2^{|I|} - (|I|+1) = 2^3 - 4 = 4 \) conditions) where some of the inequalities are satisfied but others are not. In particular, Exercise 2.6 already shows that pairwise independence may not imply mutual independence.

**Mutual independence (with \( I \) arbitrary, countable or uncountable)** The events \( \{E_i, i \in I\} \) are said to be mutually independent if for each finite subset \( J \subset I \) with \( 0 < |J| < \infty \), the events \( \{E_j, j \in J\} \) are mutually independent. It is easy to check that this definition is equivalent to the following requirement [Exercise 2.9].

**Fact 2.2.1** When \( I \) is an uncountable index set, the collection \( \{E_i, i \in I\} \) of events in \( \mathcal{F} \) are mutually independent if and only if the conditions

\[
P[\cap_{k \in K} E_k] = \prod_{k \in K} P[E_k], \quad K \subset I, |K| < \infty, |K| \geq 1
\]

are all simultaneously satisfied.
Set-theoretic operations preserve independence in the following sense.

**Theorem 2.2.1** Consider a collection \( \{E_i, i \in I\} \) of events in \( \mathcal{F} \) where \( I \) is an arbitrary index set. If the events \( \{E_i, i \in I\} \) are mutually independent, then the following statements hold:

(i) For every subset \( J \subseteq I \), the events \( \{E_j, j \in J\} \) are mutually independent.

(ii) Taking complements does not affect mutual independence: For any subset \( C \subseteq I \) (possibly empty), the events \( \{E_i, i \in C; E_j^c, j \in I - C\} \) are mutually independent.

(iii) Partitioning does not affect mutual independence: The events \( \{G_k, k \in K\} \) are mutually independent where \( K \) is an index set, \( \{I_k, k \in K\} \) is a partition of \( I \) and for each \( k \) in \( K \), the event \( G_k \) is defined by set-theoretic operations exclusively on the events \( \{E_i, i \in I_k\} \). Here set-theoretic operations refer to taking the complement of a set, union and intersection.

Part (i) is trivial, and although Part (ii) is subsumed by Part (iii), we invite the reader to provide a direct proof in Exercise 2.11. The proof of (iii) is more delicate and will not be given here.

The next two sections provide examples where the notion of independence plays a major role.

### 2.3 Modeling repeated coin tosses

In this section we discuss a class of random experiments associated with a well-known game of chance, namely the repeated throw of a coin. Historically this situation has provided much impetus for the early development of Probability Theory. It illustrates a probabilistic paradigm that recurs in many applications where the random experiment of interest consists of repeated random trials (or sub-experiments), each with exactly two possible outcomes, carried out under identical and independent conditions.

To set the stage, we first describe the random experiment of interest in some more detail: A two-sided coin is tossed repeatedly \( n \) times (with \( n \) a positive integer). Each toss results in one of two outcomes, say \( H = \) “Head” and \( T = \) “Tail” – It is often convenient to label the outcomes as \( H = 1 \) and \( T = 0 \) or even as \( H = 1 \) and \( T = -1 \). Furthermore we assume that the \( n \) successive tosses do form independent trials, each carried out under identical conditions. This implies that the likelihood of occurrence in each trial remains the same throughout the \( n \) trials, say \( p \) (resp. \( 1 - p \)) for any coin toss resulting in \( H \) (resp. \( T \)) with \( p \) a scalar in \([0, 1]\).
2.3. MODELING REPEATED COIN TOSSES

From now on we use the labeling convention \( H = 1 \) and \( T = 0 \) so that the outcome of the random experiment can be represented by a binary sequence, i.e., a sequence of 0’s and 1’s, of length \( n \). Put differently, each outcome is a word of length \( n \) with entries drawn from \( \{0, 1\} \). This leads to taking \( \Omega = \{0, 1\}^n \) for the sample space with generic element \( \omega \) given by \( \omega = (\omega_1, \ldots, \omega_n) \) where \( \omega_i \) is an element of \( \{0, 1\} \) for each \( i = 1, \ldots, n \). Following the approach in Section 1.5, with \( \mathcal{F} = \mathcal{P}(\Omega) \) as usual, we will construct the appropriate probability measure \( P \) on \( \mathcal{P}(\Omega) \) by identifying a pmf \( (p(\omega), \omega \in \{0, 1\}^n) \) which reflects the probabilistic properties described earlier.

To do so, for each \( k = 1, \ldots, n \) we define the events

\[
H_k \equiv \{ \omega \in \Omega : \omega_k = 1 \}
\]

and

\[
T_k \equiv \{ \omega \in \Omega : \omega_k = 0 \}.
\]

The event \( H_k \) (resp. \( T_k \)) contains the outcomes of the \( n \) tosses for which the \( k^{th} \) toss results in \( H \) (resp. \( T \)). The disjoint sets \( H_k \) and \( T_k \) are complementary sets in \( \Omega \) since \( H_k \cup T_k = \{0, 1\}^n \), hence \( T_k = H_k^c \). That the \( n \) successive tosses form independent trials, each carried out under identical conditions, naturally translates into the events \( \{H_k, k = 1, \ldots, n\} \) being mutually independent with

\[
P[H_k] = p = 1 - P[T_k], \quad k = 1, \ldots, n.
\]

Now pick \( \omega \) in \( \Omega \), and introduce the index sets \( H(\omega) \equiv \{k = 1, \ldots, n : \omega_k = 1\} \) and \( T(\omega) \equiv \{k = 1, \ldots, n : \omega_k = 0\} \). The disjoint index sets \( H(\omega) \) and \( T(\omega) \) are obviously complement of each other in \( \{1, \ldots, n\} \) since \( H(\omega) \cup T(\omega) = \{1, \ldots, n\} \). Noting the representation

\[
\{\omega\} = (\cap_{k \in H(\omega)} H_k) \cap (\cap_{\ell \in T(\omega)} T_\ell),
\]

we get

\[
P[\{\omega\}] = \prod_{k \in H(\omega)} P[H_k] \cdot \prod_{\ell \in T(\omega)} P[T_\ell]
\]

\[
= p^{|H(\omega)|} \cdot (1 - p)^{|T(\omega)|}
\]

upon invoking Part (ii) of Theorem 2.2.1, namely that taking complements does not change mutual independence. In the last expression, \( |H(\omega)| \) (resp. \( |T(\omega)| \)) denotes the number of trials (or equivalently, coin tosses) in the sample \( \omega \) that result in \( H \) (resp. \( T \)).
In conclusion the assumptions that the \( n \) successive tosses form independent trials, each carried out under identical conditions, lead to the pmf \( (p(\omega), \omega \in \{0, 1\}^n) \) being given by

\[
p(\omega) = p^{|H(\omega)|} \cdot (1 - p)^{|T(\omega)|},
\]

(2.9)

since \(|H(\omega)| + |T(\omega)| = n\). The case \( p = \frac{1}{2} \) is often referred to as the fair case and is explored in Exercise 2.17.

### 2.4 A probabilistic proof of a formula by Euler

The Riemann function \( \zeta : (0, \infty) \to [0, +\infty] \) is defined by

\[
\zeta(s) \equiv \sum_{m=1}^{\infty} \frac{1}{m^s}, \quad s > 0.
\]

(2.10)

It is easy to show that \( \zeta(s) = \infty \) on the range \( 0 < s \leq 1 \) and that \( \zeta(s) < \infty \) for \( s > 1 \). The following identity was established by Euler.

**Theorem 2.4.1** It holds that

\[
\zeta(s) = \prod_{p \in \mathcal{P}} \frac{1}{1 - p^{-s}}, \quad s > 1
\]

(2.11)

where \( \mathcal{P} \) denotes the set of prime numbers.

We will now provide a probabilistic proof of this remarkable identity, thereby illustrating the power of probabilistic thinking!

**Proof.** Fix \( s > 1 \). The basic idea of the proof is to construct a discrete probability model \((\Omega, \mathcal{F}, P_s)\) tailored to the Riemann function \( \zeta(s) \): Take \( \Omega = \mathbb{N}_0 \) and \( \mathcal{F} = \mathcal{P}(\mathbb{N}_0) \). Following the approach developed in Section 1.5 we define \( P_s \) on \( \mathcal{P}(\mathbb{N}_0) \) through the pmf \( (p_s(n), n = 1, 2, \ldots) \) given by

\[
p_s(n) = \frac{n^{-s}}{\zeta(s)}, \quad n = 1, 2, \ldots
\]

From the definition (2.10) of \( \zeta(s) \) this definition is well posed and the collection \((p_s(n), n = 1, 2, \ldots)\) is indeed a pmf on \( \mathbb{N}_0 \).
Next, for each \( k = 1, 2, \ldots \), we introduce the subset \( M_k \) of \( \mathbb{N}_0 \) given by \( M_k = \{ kn, n = 1, 2, \ldots \} \). An elementary calculation then shows that
\[
\mathbb{P}_s[M_k] = \sum_{n=1}^{\infty} p_s(kn) = \sum_{n=1}^{\infty} \frac{(kn)^{-s}}{\zeta(s)} = \frac{1}{k^s}.
\]

For each \( \ell = 2, 3, \ldots \), if \( p_1, \ldots, p_\ell \) are positive integers with no common divisors, then it is the case [Part (i) of Exercise 2.20] that
\[
\bigcap_{r=1}^{\ell} M_{p_r} = M_{p_1 p_2 \cdots p_\ell}.
\]
It then follows that
\[
\mathbb{P}_s \left[ \bigcap_{r=1}^{\ell} M_{p_r} \right] = \mathbb{P}_s[M_{p_1 p_2 \cdots p_\ell}]
= \frac{1}{(p_1 p_2 \cdots p_\ell)^s},
= \prod_{r=1}^{\ell} \mathbb{P}_s[M_{p_r}].
\]

As this conclusion obviously holds for any collection of prime numbers \( p_1, \ldots, p_\ell \), we have that the events \( \{M_p, p \in \mathcal{P}\} \) are mutually independent; see also Fact 2.2.1.

Now label the prime numbers by increasing size, say \( \mathcal{P} = \{ p_1, p_2, \ldots \} \) with \( p_1 < p_2 < \ldots \). By Part (ii) of Theorem 2.2.1 the mutual independence of the events \( \{M_p, p \in \mathcal{P}\} \) implies that of the complementary events \( \{M_c^c, p \in \mathcal{P}\} \). Thus, for any \( \ell = 1, 2, \ldots \), it holds that
\[
\mathbb{P}_s \left[ \bigcap_{r=1}^{\ell} M_{p_r}^c \right] = \prod_{r=1}^{\ell} \mathbb{P}_s[M_{p_r}^c] = \prod_{r=1}^{\ell} (1 - p_r^{-s}).
\]

Let \( \ell \) go to infinity in (2.14): On one hand we get
\[
\lim_{\ell \to \infty} \mathbb{P}_s \left[ \bigcap_{r=1}^{\ell} M_{p_r}^c \right] = \lim_{\ell \to \infty} \prod_{r=1}^{\ell} (1 - p_r^{-s}) = \prod_{r=1}^{\infty} (1 - p_r^{-s}).
\]

On the other hand we have \( \bigcap_{p \in \mathcal{P}} M_c^c = \{1\} \) [Part (ii) of Exercise 2.20] and
\[
\lim_{\ell \to \infty} \mathbb{P}_s \left[ \bigcap_{r=1}^{\ell} M_{p_r}^c \right] = \mathbb{P}_s[\bigcap_{r=1}^{\infty} M_{p_r}^c] = \mathbb{P}_s[\{1\}] = \zeta(s)^{-1}.
\]

The first equality expresses the continuity from above of the probability measure \( \mathbb{P}_s \) discussed in Lemma 4.1.2 (applied with \( E_\ell = \bigcap_{r=1}^{\ell} M_{p_r}^c \) for all \( \ell = 1, 2, \ldots \)). The proof of Theorem 2.4.1 is now complete as we combine (2.15) and (2.16).
2.5 Conditional probabilities

Conditional probabilities naturally arise when independence does not hold. We begin with a classical definition.

**Definition 2.5.1**

Consider an event $B$ in $\mathcal{F}$ such that $\mathbb{P}[B] > 0$. The conditional probability of the event $A$ in $\mathcal{F}$ given $B$ is defined as the ratio

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$  

(2.17)

When $\mathbb{P}[B] = 0$ it is customary to take $\mathbb{P}[A|B]$ to be arbitrary in $[0, 1]$. However, the relation

$$\mathbb{P}[A|B] \mathbb{P}[B] = \mathbb{P}[A \cap B], \quad A \in \mathcal{F}$$  

(2.18)

is always true regardless of whether $\mathbb{P}[B] > 0$ or not: When $\mathbb{P}[B] > 0$ this is clear from (2.17), while if $\mathbb{P}[B] = 0$, then $\mathbb{P}[A \cap B] = 0$ and $\mathbb{P}[A|B] \mathbb{P}[B] = 0$, irrespective of the arbitrary value selected for $\mathbb{P}[A|B]$. The following fact is an easy consequence of the definitions.

**Fact 2.5.1** With $\mathbb{P}[B] > 0$, the mapping $Q_B : \mathcal{F} \to [0, 1]$ defined by

$$Q_B(A) = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, \quad A \in \mathcal{F}$$

is a probability measure on $\mathcal{F}$.

Incidentally it is this fact that is often invoked to justify that $\mathbb{P}[\cdot|B]$ be selected as a probability measure on $\mathcal{F}$ when $\mathbb{P}[B] = 0$.

Pairwise independence can be easily characterized in terms of conditional probabilities.

**Fact 2.5.2** Let $A$ and $B$ be two events in $\mathcal{F}$. Under the condition $\mathbb{P}[B] > 0$, the events $A$ and $B$ are independent if and only if $\mathbb{P}[A|B] = \mathbb{P}[A]$.

In other words, the events $A$ and $B$ are independent if the conditional probability of $A$ given $B$ coincides with its unconditional probability $\mathbb{P}[A]$. This is a simple consequence of (2.18) and of the definition of pairwise independence; its proof is left as an exercise (Exercise 2.29).

We close this chapter with two easy, but important, consequences associated with the notion of conditional probability.
2.5. CONDITIONAL PROBABILITIES

Definition 2.5.2

With $I$ a countable index set, the events $\{B_i, \ i \in I\}$ in $\mathcal{F}$ form an $\mathcal{F}$-measurable partition of $\Omega$ whenever

$$B_i \cap B_j = \emptyset, \quad i, j \in I \quad \text{and} \quad \cup_{i \in I} B_i = \Omega.$$ 

This definition does not preclude that $\mathbb{P}[B_i] = 0$ for some $i$ in $I$. However, the second condition yields

$$\sum_{i \in I} \mathbb{P}[B_i] = 1$$ 

and therefore $\mathbb{P}[B_i] > 0$ for at least one index $i$ in $I$.

The Law of Total Probability

For each $A$ in $\mathcal{F}$, the obvious decomposition $A = \cup_{i \in I}(A \cap B_i)$ yields

$$\mathbb{P}[A] = \sum_{i \in I} \mathbb{P}[A \cap B_i]$$

(2.19) $$= \sum_{i \in I} \mathbb{P}[A|B_i] \mathbb{P}[B_i], \quad A \in \mathcal{F}.$$ 

Put differently,

$$\mathbb{P}[A] = \sum_{i \in I} Q_{B_i}(A) \mathbb{P}[B_i], \quad A \in \mathcal{F}$$

in the notation used in Fact 2.5.1.

Bayes’ rule – From prior probabilities to posterior probabilities

Consider any event $A$ in $\mathcal{F}$ such that $\mathbb{P}[A] > 0$. For each $k$ in $I$, we have

$$\mathbb{P}[B_k|A] = \frac{\mathbb{P}[B_k \cap A]}{\mathbb{P}[A]}$$

$$= \frac{\mathbb{P}[A|B_k] \mathbb{P}[B_k]}{\mathbb{P}[A]}$$

(2.20) $$= \frac{\mathbb{P}[A|B_k] \mathbb{P}[B_k]}{\sum_{i \in I} \mathbb{P}[A|B_i] \mathbb{P}[B_i]} \quad [\text{By the Law of Total Probability}]$$

This last relation, which gives the posterior probability $\mathbb{P}[B_k|A]$ in terms of the likelihoods $\{\mathbb{P}[A|B_i], \ i \in I\}$ and the prior probabilities $\{\mathbb{P}[B_i], \ i \in I\}$, is a celebrated relation known as Bayes’ rule or Bayes’ law. It plays a central role in many branches of Statistics and Data Science.
CHAPTER 2. ELEMENTARY PROBABILITY THEORY

2.6 Exercises

Unless otherwise specified a probability triple \((\Omega, \mathcal{F}, \mathbb{P})\) is assumed given. Exercises 2.1–2.5 assume the setting of Section 2.1.

Ex. 2.1 Prove Poincaré’s formulae (2.1) and (2.2) [HINT: The result is true when \(|I| = 2\), and then use an induction argument on the size of \(I\)].

Ex. 2.2 What happens to Poincaré’s formula (2.1) when the events \(\{E_i, \ i \in I\}\) are disjoint?

Ex. 2.3 Prove the bounds (2.3) and (2.4) [HINT: The results are true when \(|I| = 2\), and then use an induction argument on the size of \(I\)].

Ex. 2.4 The bounds (2.3) and (2.4) take a particularly simple form when for each \(k = 1, 2\), the individual probabilities \(\mathbb{P}\left[\bigcap_{i_k = 1}^{k} E_{j_i}\right]\) do not depend on the index set \(i_1 < \cdots < i_k\). In such situations, show that

\[
|I|\mathbb{P}[E_1] - \frac{|I||I| - 1}{2}\mathbb{P}[E_1 \cap E_2] \leq \mathbb{P}\left[\bigcup_{i \in I} E_i\right] \leq |I|\mathbb{P}[E_1].
\]

Ex. 2.5 Show that

\[
\mathbb{P}\left[\bigcap_{i \in I} E_i\right] \geq \sum_{i \in I} \mathbb{P}[E_i] - (|I| - 1).
\]

a. First proof: The result is true when \(|I| = 2\) by virtue of Exercise 1.7. Then proceed with an induction argument on the size of \(I\).

b. Second proof: Apply the union bound to the collection \(\{E_i^c, \ i \in I\}\).

Ex. 2.6 An item is selected uniformly from a set comprising four distinct objects labelled 1, 2, 3, 4. To model this experiment we take \(\Omega = \{1, 2, 3, 4\}, \mathcal{F} = \mathcal{P}(\Omega)\) and \(\mathbb{P}\) given by the uniform pmf \(p(1) = \ldots = p(4) = \frac{1}{4}\). On this probability space define three events, say \(A, B\) and \(C\), such that the events \(A, B\) and \(C\) are pairwise independent but not mutually independent, i.e., \(\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]\), \(\mathbb{P}[B \cap C] = \mathbb{P}[B]\mathbb{P}[C]\) and \(\mathbb{P}[A \cap C] = \mathbb{P}[A]\mathbb{P}[C]\) and yet \(\mathbb{P}[A \cap B \cap C] \neq \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C]\).

Ex. 2.7 An item is selected uniformly from a set comprising eight distinct objects labelled \(1, \ldots, 8\). We model this experiment by taking \(\Omega = \{1, \ldots, 8\}, \mathcal{F} = \mathcal{P}(\Omega)\) and \(\mathbb{P}\) given by the uniform pmf \(p(1) = \ldots = p(8) = \frac{1}{8}\). On this probability
space define three events, say $A$, $B$ and $C$, such that $\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]$, $\mathbb{P}[A \cap C] = \mathbb{P}[A] \mathbb{P}[C]$ and $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \mathbb{P}[B] \mathbb{P}[C]$, yet $\mathbb{P}[B \cap C] \neq \mathbb{P}[B] \mathbb{P}[C]$. In other words, the pairs of events $A$ and $B$, and $A$ and $C$ are each pairwise independent and the condition $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \mathbb{P}[B] \mathbb{P}[C]$ holds. However, the events $B$ and $C$ are not pairwise independent – This illustrates that the three events $A$, $B$ and $C$ are not mutually independent.

**Ex. 2.8** Two identical six-facetted dice are cast one after the other under identical and independent conditions, and the outcomes recorded. We model this experiment by taking $\Omega = \{(k, \ell), \ k, \ell = 1, \ldots\}, \mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}$ given by the uniform pmf $p(k, \ell) = \frac{1}{36} \ (k, \ell = 1, \ldots, 6)$. Consider the events $A$ and $B$ given by

$$A \equiv \text{[The sum of the two outcomes is 7]}$$

and

$$B \equiv \text{[The first outcome is 3]}.$$  

Show that the events $A$ and $B$ are independent. Although there is intuitively a form of “influence” between the events $A$ and $B$ – After all getting a “3” in the first outcome obviously affects the second outcome that could produce a sum “7”, these events are independent according to definition given in Section 2.2 because the realization of one event does not affect the probability of the other.

**Ex. 2.9** Prove the equivalence stated in Fact 2.2.1.

**Ex. 2.10** Let $\{E_i, \ i \in I\}$ denote a collection of events in $\mathcal{F}$.

a. With $I$ countably infinite, explain why the definition of independence requiring

$$\mathbb{P}[\bigcap_{j \in J} E_j] = \prod_{j \in J} \mathbb{P}[E_j], \quad J \subseteq I, \quad 1 \leq |J| \quad (2.21)$$

is equivalent to the one given in Section 2.2 (and equivalent to the one given in Fact 2.2.1) – However, note that when $J$ is countably infinite the conditions (2.21) are usually not informative.

b. With $I$ uncountable, explain why a definition of independence requiring

$$\mathbb{P}[\bigcap_{j \in J} E_j] = \prod_{j \in J} \mathbb{P}[E_j], \quad J \subseteq I, \quad 1 \leq |J|$$

makes no mathematical sense.
Ex. 2.11 Prove Part (ii) of Theorem 2.2.1 [HINT 1: It suffices to consider the case when $I$ is finite] [HINT 2: Proceed by induction on the size of the index set $C$ when evaluating the probabilities
\[
\mathbb{P} \left[ \left( \bigcap_{j \in C} E_j^c \right) \cap \left( \bigcap_{i \in I \setminus C} E_i \right) \right]
\]
where $C \subseteq I$. Start with the case $|C| = 1$.]

Ex. 2.12 If the events $E_1, \ldots, E_n$ in $\mathcal{F}$ are mutually independent events such that $\mathbb{P} \left[ \bigcup_{i=1}^n E_i \right] = 1$, show that $\mathbb{P} \left[ E_k \right] = 1$ for some index $k = 1, \ldots, n$. Is the index $k$ unique?

Ex. 2.13 Let $E$, $F$ and $G$ denote three events in $\mathcal{F}$ which are mutually independent, and assume that $0 < \mathbb{P} \left[ E \right], \mathbb{P} \left[ F \right] < 1$. Under what conditions are the events $E \cap G$ and $F \cap G$ independent?

Ex. 2.14 Let $A$, $E_1$ and $E_2$ denote three events in $\mathcal{F}$. Assuming that for each $k = 1, 2$, the events $A$ and $E_k$ are independent, show that the events $A$ and $E_1 \cap E_2$ are independent if and only if the events $A$ and $E_1 \cup E_2$ are independent.

Ex. 2.15 Let $A$ denote an event in $\mathcal{F}$ with $0 < \mathbb{P} \left[ A \right] < 1$ (in order to avoid trivial situations of limited interest). Define the collection $\mathcal{F}_A$ of events in $\mathcal{F}$ by
\[
\mathcal{F}_A = \{ F \in \mathcal{F} : \mathbb{P} \left[ F \cap A \right] = \mathbb{P} \left[ F \right] \cdot \mathbb{P} \left[ A \right] \}.
\]

a. Show that both $\Omega$ and the empty set $\emptyset$ belong to $\mathcal{F}_A$.

b. Show that $\mathcal{F}_A$ is closed under complementarity, i.e., if $F$ is an element of $\mathcal{F}_A$, then so is its complement $F^c$.

c. Is the family $\mathcal{F}_A$ a $\sigma$-field on $\Omega$? Prove or give a counterexample!

Ex. 2.16 Consider the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ where (i) $\Omega = \{1, \ldots, p\}$ for some prime number $p$; (ii) $\mathcal{F}$ is the power set of $\Omega$; and (iii) the probability assignment $\mathbb{P}$ is uniform in the sense that $\mathbb{P} \left[ A \right] = \frac{|A|}{p}$ for every subset $A$ of $\Omega$. Consider now two independent events $A$ and $B$, neither of which is empty. What can you say concerning these sets?

Ex. 2.17 Consider the situation discussed in Section 2.3. The pmf $(p(\omega), \omega \in \{0, 1\}^n)$ described at (2.9) was obtained by translating the requirement that the $n$ successive tosses form independent trials, each carried out under identical conditions. This was done by positing the mutual independence of the events $\{H_k, k = 1, 2, \ldots, n\}$ defined at (2.5) with probability assignment (2.7).
If the coin is fair, i.e., \( p = \frac{1}{2} \), then the expression (2.9) reduces to

\[
(2.22) \quad p(\omega) = 2^{-n}, \quad \omega \in \{0, 1\}^n.
\]

Since \( \Omega = \{0, 1\}^n \) for this model, hence \( |\Omega| = 2^n \), we conclude from the discussion in Section 1.6 that the pmf (2.22) corresponds to uniform selection.

Conversely, if we consider the uniform probability assignment (2.22), show that the events \( \{H_k, \ k = 1, 2, \ldots, n\} \) defined at (2.5) are necessarily mutually independent.

**Ex. 2.18** The situation discussed in Exercise 2.17 can be further generalized: Consider \( n \) probability triples, each with a finite sample space, say \((\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1), \ldots, (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)\). Consider the measurable space \((\Omega, \mathcal{F})\) where \( \Omega \equiv \Omega_1 \times \ldots \times \Omega_n \) and \( \mathcal{F} \equiv \mathcal{P}(\Omega) \).

a. Show that there exists a unique probability measure \( \mathbb{P} \) on \( \mathcal{P}(\Omega) \) such that

\[
\mathbb{P}[E_1 \times \ldots \times E_n] = \prod_{k=1}^{n} \mathbb{P}_k[E_k], \quad E_k \in \mathcal{P}(\Omega_k), \quad k = 1, \ldots, n
\]

and give an expression for the probabilities

\[
\mathbb{P}[E], \quad E \in \mathcal{P}(\Omega).
\]

[**HINT:** What is the value of \( \mathbb{P}[\{\omega\}] \) for each \( \omega \) in \( \Omega \)?]

b. Show that the probability measure \( \mathbb{P} \) is uniform on \( \mathcal{P}(\Omega) \) if and only if each probability measure \( \mathbb{P}_k \) is uniform on \( \mathcal{P}(\Omega_k), k = 1, \ldots, n \).

**Ex. 2.19** A fair coin is rolled \( n \) times under identical and independent conditions, as in Exercise 2.17. We adopt the model discussed in Section 2.3 (with \( p = \frac{1}{2} \)).

With distinct \( i, j = 1, \ldots, n \), define the event \( E_{ij} \) as the event where the outcomes of the \( i^{th} \) and \( j^{th} \) tosses are identical (e.g., both are heads). Show that the \( \frac{n(n-1)}{2} \) events \( \{E_{ij}, 1 \leq i < j \leq n\} \) are pairwise independent but not mutually independent!

**Ex. 2.20** In the proof of Theorem 2.4.1:

(i) Prove the set equality (2.20) [**HINT:** Prove it first for \( \ell = 2 \), and use induction on \( \ell \) to establish the general case].

(ii) Show that \( \cap_{p \in \mathcal{P}} M^\ell_p = \{1\} \).

**Ex. 2.21** Establish Fact 2.5.1

**Ex. 2.22** In Definition 2.5.1 show through examples that both inequalities \( \mathbb{P}[A] < \mathbb{P}[A|B] \) and \( \mathbb{P}[A|B] < \mathbb{P}[A] \) are possible – Assume that \( \mathbb{P}[B] > 0 \).
Ex. 2.23 Given are three scalars $\alpha, \beta$ and $\gamma$ in $(0, 1)$. Construct a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ and a pair of events $E$ and $F$ in $\mathcal{F}$ such that $\mathbb{P}[F] = \beta$, $\mathbb{P}[E|F] = \alpha$ and $\mathbb{P}[E|F^c] = \gamma$.

Ex. 2.24 Let $E$ and $F$ be events in $\mathcal{F}$ with $\mathbb{P}[E] > 0$ and $\mathbb{P}[F] > 0$. We say that event $E$ is positively correlated with event $F$ if $\mathbb{P}[E|F] \geq \mathbb{P}[E]$.

a. Show the equivalence of the following three statements (i)-(iii) below:

(i) Event $E$ is positively correlated with event $F$

(ii) Event $F$ is positively correlated with event $E$

(iii) Event $E^c$ is positively correlated with event $F^c$.

b. Construct a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ and three events $E$, $F$ and $G$ in $\mathcal{F}$ such that $\mathbb{P}[E|F] > \mathbb{P}[E]$ [Event $E$ is (strictly) positively correlated with event $F$] but $\mathbb{P}[E|G] > \mathbb{P}[E]$ [Event $E$ is not positively correlated with event $G$]. [HINT: Take $\Omega = \{1, 2, 3\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and the uniform probability assignment on $\mathcal{F}$.]

Ex. 2.25 The decomposition formula for conditional probabilities: Given three events $E$, $F$ and $G$ in $\mathcal{F}$ such that $\mathbb{P}[F \cap G] > 0$ and $\mathbb{P}[F \cap G^c] > 0$, show that

$$\mathbb{P}[E|F] = \mathbb{P}[G|F] \mathbb{P}[E|F \cup G] + \mathbb{P}[G^c|F] \mathbb{P}[E|F \cup G^c].$$

Ex. 2.26 Consider events $E_1, \ldots, E_n$ in $\mathcal{F}$ that are disjoint with $\mathbb{P}[E_i] > 0$ for all $i = 1, \ldots, n$. For any event $E$ in $\mathcal{F}$ show that the bounds

$$\min_{i=1,\ldots,n} \mathbb{P}[F|E_i] \leq \mathbb{P}[E|\bigcup_i E_i] \leq \max_{i=1,\ldots,n} \mathbb{P}[E|E_i]$$

hold. [HINT: Let $\alpha_1, \ldots, \alpha_n$ be non-negative scalars such that $\alpha_1 + \ldots + \alpha_n = 1$. Then, for any $x_1, \ldots, x_n$ in $\mathbb{R}$ the inequalities

$$\min_{i=1,\ldots,n} x_i \leq \sum_{i=1}^n \alpha_i x_i \leq \max_{i=1,\ldots,n} x_i$$

are satisfied].

Ex. 2.27 We return to Exercise 1.9: Let $S$ denote a finite set, say $S = \{1, \ldots, n\}$ for some positive integer $n$. The random experiment $\mathcal{E}$ consists in selecting uniformly an ordered pair $A$ and $B$ of (possibly empty) subsets of $S$. Using the model developed there, evaluate $\mathbb{P}[\mathcal{I}|B]$, $B \subseteq S$, and then using the Law of Total Probability.
2.6. EXERCISES

Ex. 2.28 The following fact is useful when modeling situations associated with sequential decision making: Given a probability model \((\Omega, \mathcal{F}, \mathbb{P})\), with events \(A_1, \ldots, A_n\) in \(\mathcal{F}\), show that

\[
\mathbb{P}[A_1 \cap \ldots \cap A_n] = \mathbb{P}[A_1] \cdot \prod_{i=2}^{n} \mathbb{P}[A_i | A_1 \cap \ldots \cap A_{i-1}].
\]

[HINT: The observation

\[
\mathbb{P}[A_1 \cap \ldots \cap A_n] = \mathbb{P}[A_n | A_1 \cap \ldots \cap A_{n-1}] \cdot \mathbb{P}[A_1 \cap \ldots \cap A_{n-1}]
\]

suggests a proof by induction on \(n\).

Ex. 2.29 Establish Fact 2.5.2

We close with several urn problems

Ex. 2.30 Consider two urns, say \(U_1\) and \(U_2\), each of which contains contains \(R\) red balls and \(B\) blue balls. A ball is drawn at random from urn \(U_1\), and put in urn \(U_2\). Urn \(U_2\) is then well stirred and shaken, and a ball is drawn at random from urn \(U_2\).

a. Describe a complete probability model \((\Omega, \mathcal{F}, \mathbb{P})\) to model this situation.
b. Compute the probability that the ball drawn from urn \(U_2\) is red.

Ex. 2.31 There are three urns, say \(U_1, U_2\) and \(U_3\). Urn \(U_1\) contains \(R_1\) red balls and \(B_1\) blue balls, urn \(U_2\) contains \(R_2\) red balls and \(B_2\) blue balls, and urn \(U_3\) contains \(R_3\) red balls and \(B_3\) blue balls. An urn is selected at random, and then a ball is selected at random from it.

a. Describe a complete probability model \((\Omega, \mathcal{F}, \mathbb{P})\) to model this situation.
b. Compute the probability that the selected ball came from urn \(U_1\) if the ball selected is red.

Ex. 2.32 Consider \(N\) urns (with \(N \geq 2\), say \(U_1, \ldots, U_N\), each of which initially contains \(R\) red balls and \(B\) blue balls. Each of the urns has been well stirred and shaken! A ball is drawn at random from urn \(U_1\), and put in urn \(U_2\) which is then well stirred and shaken! Then, a ball is drawn at random from urn \(U_2\) and put in urn \(U_3\). The process is repeated until a ball is finally drawn at random from last urn \(U_N\).

a. Describe a complete probability model \((\Omega, \mathcal{F}, \mathbb{P})\) to model this situation.
b. Compute the probability that the ball selected from urn \(U_1\) is red.

Ex. 2.33
Chapter 3

Measurable mappings:
A tale of \( \sigma \)-fields

Many situations of interest naturally require that the sample space be \textit{uncountable} – For instance, for some models it is appropriate for the sample space to be \( \mathbb{R}^p \) (or a subset thereof); see Chapter ?? for some concrete examples. Unfortunately in such cases determining the appropriate \( \sigma \)-field of events on which to define the probability measures is technically more delicate: In a nutshell the required \( \sigma \)-additivity imposes too many constraints if the probability measure is to be defined on the entire power set of the sample space; this precludes that the power set of the sample space be used as the \( \sigma \)-field (as we did for the countable case in Section 1.5).

In the present chapter we start to address the challenge of constructing \( \sigma \)-fields on uncountable sample spaces such as (subsets of) \( \mathbb{R}^p \). The basic idea is to tie the definition to the underlying topological properties of the sample space. This arises from the need to assign a likelihood of occurrence to certain subsets of the sample space. In \( \mathbb{R}^p \) this leads to introducing the important notion of the Borel \( \sigma \)-field on \( \mathbb{R}^p \) and the related concept of Borel measurability. The narrative continues in Chapter ?? where some intuition behind these definitions is developed; there we introduce Carathéodory’s Extension Theorem and explain its use in constructing measures in the context of some important applications.

3.1 Measurable mappings

We begin by discussing the notion of \textit{measurability} of a mapping: To fix the notation, let \( S_a \) and \( S_b \) be arbitrary sets (but possibly identical). For any mapping
$g : S_a \to S_b$, recall that

$$g^{-1}(E_b) \equiv \{ s_a \in S_a : g(s_a) \in E_b \}, \quad E_b \in \mathcal{P}(S_b).$$

With $\mathcal{H}_b$ denoting a collection of subsets of $S_b$, it will be natural to extend this notation to collections of subsets by writing

$$g^{-1}(\mathcal{H}_b) \equiv \{ g^{-1}(E_b) : E_b \in \mathcal{H}_b \}.$$

Consider now the situation where the sets $S_a$ and $S_b$ are equipped with $\sigma$-fields $\mathcal{S}_a$ and $\mathcal{S}_b$, respectively – Thus, the pairs $(S_a, \mathcal{S}_a)$ and $(S_b, \mathcal{S}_b)$ are measurable spaces. In cases where $S_a = S_b \equiv S$ we could in principle have distinct $\sigma$-fields $\mathcal{S}_a$ and $\mathcal{S}_b$ on $S$.

**Definition 3.1.1**

When the sets $S_a$ and $S_b$ are equipped with $\sigma$-fields $\mathcal{S}_a$ and $\mathcal{S}_b$, respectively, the mapping $g : S_a \to S_b$ is said to be $(\mathcal{S}_a, \mathcal{S}_b)$-measurable if the conditions

$$g^{-1}(E_b) \in \mathcal{S}_a, \quad E_b \in \mathcal{S}_b \quad (3.1)$$

all hold.

The conditions (3.1) can be rewritten more compactly as

$$g^{-1}(\mathcal{S}_b) \subseteq \mathcal{S}_a. \quad (3.2)$$

Adding a third set $S_c$ equipped with a $\sigma$-field $\mathcal{S}_c$, we consider now a situation where there are now three measurable spaces $(S_a, \mathcal{S}_a), (S_b, \mathcal{S}_b)$ and $(S_c, \mathcal{S}_c)$. With mappings $g : S_a \to S_b$ and $h : S_b \to S_c$, we associate the composition mapping $h \circ g : S_a \to S_c$ given by

$$(h \circ g)(s_a) \equiv h(g(s_a)), \quad s_a \in S_a.$$

**Fact 3.1.1** If the mapping $g : S_a \to S_b$ is $(\mathcal{S}_a, \mathcal{S}_b)$-measurable and the mapping $h : S_b \to S_c$ is $(\mathcal{S}_b, \mathcal{S}_c)$-measurable, then the composition mapping $h \circ g : S_a \to S_c$ is itself $(\mathcal{S}_a, \mathcal{S}_c)$-measurable.

**Proof.** The conclusion follows from the elementary set-theoretic fact

$$\quad (h \circ g)^{-1}(E_c) = g^{-1}(h^{-1}(E_c)), \quad E_c \in \mathcal{P}(S_c) \quad (3.3)$$
when coupled with the \((S_b, S_c)\)-measurability of \(h\) and the \((S_a, S_b)\)-measurability of \(g\). Details are left to the interested reader [Exercise 3.2].

Lemma 3.1.1 given next is key to showing that the measurability of a mapping can often be determined by checking a reduced set of conditions.

**Lemma 3.1.1** Let \(\mathcal{H}_b\) be a collection of subsets of \(S_b\). For any mapping \(g : S_a \rightarrow S_b\), the following statements hold:

(i) If \(\mathcal{H}_b\) is a \(\sigma\)-field on \(S_b\), then the collection \(g^{-1}(\mathcal{H}_b)\) is a \(\sigma\)-field on \(S_a\);

(ii) More generally, we always have

\[
g^{-1}(\sigma(\mathcal{H}_b)) = \sigma(g^{-1}(\mathcal{H}_b)).
\]  

**Proof.** Claim (i): We leave it as an exercise [Exercise 3.1] to check that the collection \(g^{-1}(\mathcal{H}_b)\) is a \(\sigma\)-field on \(S_a\) when \(\mathcal{H}_b\) is a \(\sigma\)-field on \(S_b\).

Claim (ii): We now turn to establishing (3.4): By Part (i) applied to the \(\sigma\)-field \(\sigma(\mathcal{H}_b)\), the collection \(g^{-1}(\sigma(\mathcal{H}_b))\) is a \(\sigma\)-field which contains \(g^{-1}(\mathcal{H}_b)\), and the inclusion \(\sigma(g^{-1}(\mathcal{H}_b)) \subseteq g^{-1}(\sigma(\mathcal{H}_b))\) is straightforward by virtue of Fact 1.7.2.

To establish the reverse inclusion \(g^{-1}(\sigma(\mathcal{H}_b)) \subseteq \sigma(g^{-1}(\mathcal{H}_b))\), consider the collection \(\mathcal{H}_{b,g}^*\) of subsets given by

\[
\mathcal{H}_{b,g}^* \equiv \{ E_b \in \mathcal{P}(S_b) : g^{-1}(E_b) \in \sigma(g^{-1}(\mathcal{H}_b)) \}.
\]

It is easy to check that \(\mathcal{H}_{b,g}^*\) is a \(\sigma\)-field on \(S_b\) [Exercise 3.1] containing \(\mathcal{H}_b\). Therefore, \(\mathcal{H}_{b,g}^*\) contains \(\sigma(\mathcal{H}_b)\) and we get

\[
g^{-1}(\sigma(\mathcal{H}_b)) \subseteq g^{-1}(\mathcal{H}_{b,g}^*) \subseteq \sigma(g^{-1}(\mathcal{H}_b))
\]

where the last inclusion follows by the definition of \(\mathcal{H}_{b,g}^*\). This completes the proof of (3.4).

We now use Lemma 3.1.1 to obtain an equivalent definition of measurability.

**Lemma 3.1.2** If the \(\sigma\)-field \(S_b\) on \(S_a\) is generated by the collection \(\mathcal{H}_b\) of subsets of \(S_b\), i.e., \(S_b = \sigma(\mathcal{H}_b)\), then the mapping \(g : S_a \rightarrow S_b\) is \((S_a, S_b)\)-measurable if and only if the conditions

\[
g^{-1}(E_b) \in S_a, \quad E_b \in \mathcal{H}_b
\]

all hold.
In the same way that the conditions (3.2) are equivalent to (3.1), we can write the conditions (3.6) in the equivalent form

\[(3.7) \quad g^{-1}(\mathcal{H}_b) \subseteq \mathcal{S}_a. \]

The equivalence stated in Lemma 3.1.2 is operationally useful in that only the reduced subset (3.6) of conditions (associated with the generator \(\mathcal{H}_b\) for \(\mathcal{S}_b\)) needs to be checked instead of the entire set (3.1) – An important example will be discussed shortly in the next section.

**Proof.** The condition (3.2) obviously implies (3.7) since \(\mathcal{H}_b \subseteq \mathcal{S}_b\). Conversely, assume that the mapping \(g : \mathcal{S}_a \to \mathcal{S}_b\) satisfies (3.7): The equality \(g^{-1}(\mathcal{S}_b) = g^{-1}(\sigma(\mathcal{H}_b))\) obviously holds since \(\mathcal{S}_b = \sigma(\mathcal{H}_b)\) by assumption, while the equality \(g^{-1}(\sigma(\mathcal{H}_b)) = \sigma(g^{-1}(\mathcal{H}_b))\) follows from Lemma 3.1.1 – Combining these two equalities gives \(g^{-1}(\mathcal{S}_b) = \sigma(g^{-1}(\mathcal{H}_b))\). Finally, under condition (3.7) we conclude that \(\sigma(g^{-1}(\mathcal{H}_b)) \subseteq \mathcal{S}_a\) as we use the fact that \(\mathcal{S}_a\) is itself a \(\sigma\)-field; see Fact 1.7.2. This complete the proof of Lemma 3.1.2.

## 3.2 The Borel \(\sigma\)-field on \(\mathbb{R}\)

We refer to a subset \(I\) of \(\mathbb{R}\) of the form \((a, b)\) (with \(a \leq b\) in \(\mathbb{R}\)) as a bounded open interval. Let \(\mathcal{I}(\mathbb{R})\) denote the collection of all bounded open intervals of \(\mathbb{R}\).

As can be seen from the discussion in Section ??, it is quite natural to consider assigning a measure to such intervals. This requires at minimum that we consider the \(\sigma\)-field generated by \(\mathcal{I}(\mathbb{R})\) as we do next.

**Definition 3.2.1**

The Borel \(\sigma\)-field on \(\mathbb{R}\), denoted \(\mathcal{B}(\mathbb{R})\), is the smallest \(\sigma\)-field on \(\mathbb{R}\) containing all bounded open intervals of \(\mathbb{R}\), i.e., \(\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}(\mathbb{R}))\).

The Borel \(\sigma\)-field \(\mathcal{B}(\mathbb{R})\) can be generated in many different ways. To see this consider the following collections of subsets of \(\mathbb{R}\):

The bounded open intervals

\[\mathcal{H}_0(\mathbb{R}) \equiv \left\{(a, b), \quad a \leq b, \quad a, b \in \mathbb{R}\right\} = \mathcal{I}(\mathbb{R}).\]

The bounded closed intervals

\[\mathcal{H}_1(\mathbb{R}) \equiv \left\{[a, b], \quad a \leq b, \quad a, b \in \mathbb{R}\right\}.\]
The bounded open-closed intervals
\[ H_2 (\mathbb{R}) \equiv \left\{ (a, b), \quad a \leq b \right\} \quad a, b \in \mathbb{R} \].

The bounded closed-open intervals
\[ H_3 (\mathbb{R}) \equiv \left\{ [a, b), \quad a \leq b \right\} \quad a, b \in \mathbb{R} \].

The open semi-intervals
\[ H_4 (\mathbb{R}) \equiv \left\{ (-\infty, a), \quad a \in \mathbb{R} \right\} \].

The closed semi-intervals
\[ H_5 (\mathbb{R}) \equiv \left\{ [-\infty, a), \quad a \in \mathbb{R} \right\} \].

A key observation is contained in the following result.

**Lemma 3.2.1** With the notation above, it holds that
\[ \mathcal{B} (\mathbb{R}) = \sigma (H_k (\mathbb{R})) \quad k = 0, 1, \ldots, 5. \] (3.8)

The approximation arguments given in the proof of Lemma 3.2.1 can also be used in the multi-dimensional setting of Section 3.4.

**Proof.** Fix \( a \) and \( b \) in \( \mathbb{R} \) with \( a \leq b \). The set-theoretic facts \([a, b) = \cap_{n=1}^{\infty} (a - \frac{1}{n}, b + \frac{1}{n}) \) and \([a, b] = \cap_{n=1}^{\infty} (a, b + \frac{1}{n}) \) readily imply \( H_1 \subseteq \sigma (H_0) \) and \( H_2 \subseteq \sigma (H_0) \). On the other hand we also have
\[ (a, b) = \cup_{n=n(a,b)}^{\infty} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right] \quad (a, b) = \cup_{n=n(a,b)}^{\infty} \left( a, b - \frac{1}{n} \right] \]
with \( n(a, b) = \lceil 2(b - a)^{-1} \rceil \). These two equalities readily imply \( H_0 \subseteq \sigma (H_1) \) and \( H_0 \subseteq \sigma (H_2) \), respectively. It immediately follows that \( \sigma (H_0) = \sigma (H_1) \) and \( \sigma (H_0) = \sigma (H_2) \). A similar argument also shows that \( \sigma (H_0) = \sigma (H_3) \), and we conclude that \( \mathcal{B} (\mathbb{R}) = \sigma (H_k) \) for \( k = 0, 1, 2, 3 \).

In the same vein, upon noting that \( (-\infty, a] = \cap_{n=1}^{\infty} (-\infty, a + \frac{1}{n}) \) and \( (-\infty, a) = \cup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}] \), we conclude that \( H_5 \subseteq \sigma (H_4) \) and \( H_4 \subseteq \sigma (H_5) \), respectively, and the equality \( \sigma (H_4) = \sigma (H_5) \) follows.
Next, the inclusion $\mathcal{H}_4 \subseteq \sigma(\mathcal{H}_2)$ holds since $(-\infty, a] = \bigcup_{n=0}^{\infty}(a - (n + 1), a - n]$, hence $\sigma(\mathcal{H}_4) \subseteq \sigma(\mathcal{H}_2) = \mathcal{B}(\mathbb{R})$. To get the reverse inclusion, start with the observation that $(a, b) = (-\infty, b) \cap (-\infty, a]^c$. This shows that $\mathcal{H}_0 \subseteq \sigma(\mathcal{H}_4) = \sigma(\mathcal{H}_5)$, hence $\sigma(\mathcal{H}_0) = \mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{H}_4)$. The equality $\sigma(\mathcal{H}_0) = \sigma(\mathcal{H}_4)$ follows, and the proof of (3.8) for $k = 4, 5$ is complete.

In spite of its seemingly simple definition, namely as $\mathcal{B}(\mathbb{R}) \equiv \sigma(\mathcal{I}(\mathbb{R}))$, the Borel $\sigma$-field constitutes an extremely large and rather unwieldy collection of objects. To make this point even more apparent we now provide a characterization of $\mathcal{B}(\mathbb{R})$ in terms of a generator much larger than the ones appearing in Lemma 3.2.1.

We set the stage with a well-known definition from the standard topology on $\mathbb{R}$:

**Definition 3.2.2**

A subset $U$ of $\mathbb{R}$ is **open** if for every $x$ in $U$, there exists a bounded open interval $I_x$ (in $\mathcal{I}(\mathbb{R})$) containing $x$ (i.e., $x \in I_x$) and contained in $U$ (i.e., $I_x \subseteq U$). A set $F$ is said to be **closed** if its complement $F^c$ (in $\mathbb{R}$) is open.

Let $\mathcal{O}(\mathbb{R})$ denote the collection of all open subsets of $\mathbb{R}$. It is elementary to check that bounded open intervals and all open semi-intervals in $\mathcal{H}_4(\mathbb{R})$ are open sets, and that bounded closed intervals in $\mathcal{H}_1(\mathbb{R})$ and all closed semi-intervals in $\mathcal{H}_5(\mathbb{R})$ are closed sets. However the bounded open-closed intervals in $\mathcal{H}_2(\mathbb{R})$ and closed-open intervals in $\mathcal{H}_3(\mathbb{R})$ are neither open nor closed.

The key technical point that highlights the importance of bounded open intervals as building blocks for the usual topology on $\mathbb{R}$ is given next; see [?] for a proof.

**Fact 3.2.1** Any open subset $U$ in $\mathbb{R}$ can be expressed as the union of a countable collection of non-overlapping open intervals, i.e., there exists a countable collection $\{J_i, i \in I\}$ of open intervals of $\mathbb{R}$ such that

$$U = \bigcup_{i \in I} J_i \quad \text{with} \quad J_k \cap J_{\ell} = \emptyset, \quad k \neq \ell, \quad k, \ell \in I. \quad (3.9)$$

Fact 3.2.1 leads easily to the following characterization of $\mathcal{B}(\mathbb{R})$ in terms of open sets.

**Lemma 3.2.2** The smallest $\sigma$-field on $\mathbb{R}$ containing all open subsets of $\mathbb{R}$ coincides with the Borel $\sigma$-field $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$, i.e., $\mathcal{B}(\mathbb{R}) \equiv \sigma(\mathcal{O}(\mathbb{R}))$. 
3.3. CARTESIAN PRODUCTS

Proof. As pointed earlier, we already have $I(\mathbb{R}) \subseteq O(\mathbb{R})$, hence $\sigma(I(\mathbb{R})) \subseteq \sigma(O(\mathbb{R}))$, and the inclusion $B(\mathbb{R}) \subseteq \sigma(O(\mathbb{R}))$ holds. To obtain the reverse inclusion, note that $O(\mathbb{R}) \subseteq \sigma(I(\mathbb{R}))$ by Fact 3.2.1, hence $\sigma(O(\mathbb{R})) \subseteq \sigma(I(\mathbb{R}))$. In other words, $\sigma(O(\mathbb{R})) \subseteq B(\mathbb{R})$, and the proof is complete.

In short, the collection of all open sets on $\mathbb{R}$ generates the Borel $\sigma$-field $B(\mathbb{R})$, thereby highlighting its connection with the standard topology on $\mathbb{R}$.

3.3 Cartesian products

Before discussing the multi-dimensional case we introduce some facts concerning cartesian products: Let $S_a$ and $S_b$ be two arbitrary sets (possibly identical). If $H_a$ and $H_b$ are collections of subsets of $S_a$ and $S_b$, respectively, it is natural to set

$$H_a \times H_b \equiv \{ E_a \times E_b : E_a \in H_a, E_b \in H_b \}.$$ 

A set $E_a \times E_b$ in $H_a \times H_b$ is called a rectangle with sides $E_a$ in $S_a$ and $E_b$ in $S_b$. These notions generalize to more than two factors in an obvious manner.

In general the collection $H_a \times H_b$ is not a $\sigma$-field (resp. a field) on the Cartesian product $S_a \times S_b$ even if each of the collections $H_a$ and $H_b$ is itself a $\sigma$-field (resp. a field) [Exercise 3.4]. The next result shows how generators on the individual factors give rise to a natural notion of measurability on Cartesian products.

**Lemma 3.3.1** Let $S_a$ and $S_b$ be two arbitrary sets. If $H_a$ and $H_b$ are collections of subsets of $S_a$ and $S_b$, respectively, then it holds that

$$(3.10) \quad \sigma(H_a \times H_b) = \sigma(\sigma(H_a) \times \sigma(H_b)).$$

It is customary to write

$$\sigma(H_a \times H_b) = \sigma(H_a) \otimes \sigma(H_b).$$

**Proof.** As the inclusion $H_a \times H_b \subseteq \sigma(H_a) \times \sigma(H_b)$ obviously holds, we immediately get the inclusion $\sigma(H_a \times H_b) \subseteq \sigma(\sigma(H_a) \times \sigma(H_b))$. To establish the reverse inclusion

$$(3.11) \quad \sigma(\sigma(H_a) \times \sigma(H_b)) \subseteq \sigma(H_a \times H_b),$$
we proceed as follows: Define the collections
\[ \mathcal{H}^*_a \equiv \{ E_a \in \mathcal{P}(S_a) : E_a \times S_b \in \sigma(\mathcal{H}_a \times \mathcal{H}_b) \} \]
and
\[ \mathcal{H}^*_b \equiv \{ E_b \in \mathcal{P}(S_b) : S_a \times E_b \in \sigma(\mathcal{H}_a \times \mathcal{H}_b) \} . \]

It is a simple matter to check that \( \mathcal{H}^*_a \) and \( \mathcal{H}^*_b \) are \( \sigma \)-fields on \( S_a \) and \( S_b \), respectively [Exercise 3.5].

Pick an arbitrary subset \( E \) of \( S_a \times S_b \) that belongs to \( \mathcal{H}^*_a \times \mathcal{H}^*_b \). Thus, \( E = E_a \times E_b \) with \( E_a \) in \( \mathcal{H}^*_a \) and \( E_b \) in \( \mathcal{H}^*_b \). By definition both cartesian products \( E_a \times S_b \) and \( S_a \times E_b \) belong to the \( \sigma \)-field \( \sigma(\mathcal{H}_a \times \mathcal{H}_b) \), hence their intersection also belongs to \( \sigma(\mathcal{H}_a \times \mathcal{H}_b) \). However, as we note that \( (E_a \times S_b) \cap (S_a \times E_b) = E_a \times E_b \), we conclude that \( E \) is also an element of \( \sigma(\mathcal{H}_a \times \mathcal{H}_b) \). Put differently, we have just shown that
\[ \mathcal{H}^*_a \times \mathcal{H}^*_b \subseteq \sigma(\mathcal{H}_a \times \mathcal{H}_b) . \]

Obviously \( \mathcal{H}_a \subseteq \mathcal{H}^*_a \), hence \( \sigma(\mathcal{H}_a) \subseteq \mathcal{H}^*_a \) since \( \mathcal{H}^*_a \) is \( \sigma \)-field on \( S_a \) [Exercise 3.5]. We similarly have \( \sigma(\mathcal{H}_b) \subseteq \mathcal{H}^*_b \). It follows from (3.14) that
\[ \sigma(\mathcal{H}_a) \times \sigma(\mathcal{H}_b) \subseteq \sigma(\mathcal{H}_a \times \mathcal{H}_b) . \]
and the conclusion (3.11) follows. The proof of (3.10) is now complete.

\[ \boxed{\text{3.4 The Borel } \sigma \text{-fields on } \mathbb{R}^p (p = 2, 3, \ldots) } \]

As we turn to the multi-dimensional case, let \( p \) denote a fixed positive integer. In higher dimensions it also quite natural to consider assigning a measure to certain subsets of \( \mathbb{R}^p \), say subsets of the form
\[ B_1 \times \ldots \times B_p, \]
where for each \( k = 1, \ldots, p \), the set \( B_k \) is a subset in one of the collections \( \mathcal{H}_0(\mathbb{R}), \ldots, \mathcal{H}_5(\mathbb{R}) \) introduced in Section 3.2. Specializing this definition to \( \mathcal{H}_0(\mathbb{R}) \) we get the following definitions.

\textbf{Definition 3.4.1}

An bounded open rectangle \( R \) in \( \mathbb{R}^p \) is a product set of the form \( I_1 \times \ldots \times I_p \) where for each \( k = 1, \ldots, p \), the factor set \( I_k \) is a bounded open interval \( (a_k, b_k) \) (with \( a_k \leq b_k \) in \( \mathbb{R} \)). In other words,
\[ R = \{(y_1, \ldots, y_p) \in \mathbb{R}^p : a_k < y_k < b_k, \ k = 1, 2, \ldots, p \} . \]
Let $\mathcal{R}_{\text{Open}}(\mathbb{R}^p)$ denote the collection of all bounded open rectangles in $\mathbb{R}^p$. Obviously we have $\mathcal{R}_{\text{Open}}(\mathbb{R}^p) = \mathcal{H}_0(\mathbb{R}) \times \ldots \times \mathcal{H}_0(\mathbb{R}) = \mathcal{H}_0(\mathbb{R})^p$, or in a slightly different notation, $\mathcal{R}_{\text{Open}}(\mathbb{R}^p) = \mathcal{I}(\mathbb{R}) \times \ldots \times \mathcal{I}(\mathbb{R}) = \mathcal{I}(\mathbb{R})^p$, the latter clearly showing that $\mathcal{R}_{\text{Open}}(\mathbb{R}^p)$ is a natural multi-dimensional generalization of $\mathcal{I}(\mathbb{R})$.

In analogy with Definition 3.2.1 given for $p = 1$ we now introduce the notion of a Borel $\sigma$-field on $\mathbb{R}^p$.

**Definition 3.4.2**

The Borel $\sigma$-field on $\mathbb{R}^p$, denoted $\mathcal{B}(\mathbb{R}^p)$, is the smallest $\sigma$-field on $\mathbb{R}^p$ containing all bounded open rectangles in $\mathbb{R}^p$, i.e., $\mathcal{B}(\mathbb{R}^p) \equiv \sigma(\mathcal{R}_{\text{Open}}(\mathbb{R}^p))$.

Not too surprisingly, the Borel $\sigma$-field $\mathcal{B}(\mathbb{R}^p)$ is related to the usual topology on $\mathbb{R}^p$ (as the Borel $\sigma$-field $\mathcal{B}(\mathbb{R})$ was to the usual topology on $\mathbb{R}$). To clarify this connection further consider next the usual notion of an open set in $\mathbb{R}^p$ whose definition is analogous to Definition 3.2.1 given for $p = 1$.

**Definition 3.4.3**

A subset $U$ of $\mathbb{R}^p$ is open if for every $x = (x_1, \ldots, x_p)$ in $U$, there exists a bounded open rectangle $R_x$ (in $\mathcal{R}_{\text{Open}}(\mathbb{R}^p)$) containing $x$ (i.e., $x \in R_x$) and contained in $U$ (i.e., $R_x \subseteq U$). A set $F$ is said to be closed if its complement $F^c$ (in $\mathbb{R}^p$) is open.

In the notation of Definition 3.4.1, the set $U$ is open in $\mathbb{R}^p$ if for every $x = (x_1, \ldots, x_p)$ in $U$ and each $k = 1, \ldots, p$, there exist scalars $a_k(x)$ and $b_k(x)$ such that $a_k(x) < x_k < b_k(x)$ and

$$\{ (y_1, \ldots, y_p) \in \mathbb{R}^p : a_k < y_k < b_k, \ k = 1, \ldots \} \subseteq U.$$ 

Let $\mathcal{O}(\mathbb{R}^p)$ denote the collection of all open sets in $\mathbb{R}^p$.

In the scalar case, according to Fact 3.2.1 the bounded open intervals are the building blocks of the standard topology on $\mathbb{R}$. A similar situation holds in higher dimensions in that open rectangles in $\mathbb{R}^p$ are now the building blocks of the standard topology on $\mathbb{R}^p$. This is the message of the following well-known fact from topology [?].

**Fact 3.4.1** For any open set $U$ in $\mathbb{R}^p$ there exists a countable family of bounded open rectangles $\{R_i, i \in I\}$ in $\mathcal{R}_{\text{Open}}(\mathbb{R}^p)$ with countable $I$ such that $U = \bigcup_{i \in I} R_i$. 
Fact 3.4.1 easily leads to a characterization of $B(\mathbb{R}^p)$ in terms of the usual topology on $\mathbb{R}^p$; the proof, left as an easy exercise, mimics that of Lemma 3.2.2 (with $\mathcal{I}(\mathbb{R})$ replaced by $\mathcal{R}_{\text{Open}}(\mathbb{R}^p)$ and leveraging this time Fact 3.4.1) [Exercise 3.6].

**Lemma 3.4.1** The smallest $\sigma$-field on $\mathbb{R}^p$ containing all open subsets of $\mathbb{R}^p$ coincides with the Borel $\sigma$-field $B(\mathbb{R}^p)$ on $\mathbb{R}^p$, i.e., $B(\mathbb{R}^p) \equiv \sigma(\mathcal{O}(\mathbb{R}^p))$.

As in the scalar case discussed in Section 3.2 the Borel $\sigma$-field $B(\mathbb{R}^p)$ can be generated in many different ways; we leave it to the reader to explore the appropriate multi-dimensional generalization of Lemma 3.2.1. However, for reasons that will soon become apparent in later chapters, there is one generating family that occupies a central place in developing the notion of probability distribution functions [Chapter 5]: Let $\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p)$ denote the collection of all closed southwest rectangles, i.e.,

$$\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p) \equiv \left\{ J_1 \times \ldots \times J_p, \quad J_k = (-\infty, a_k] \right\}, \quad a_k \in \mathbb{R}, \quad k = 1, \ldots, p, \quad$$

This family is the $p$-dimensional analog of the one-dimensional family $\mathcal{H}_1(\mathbb{R})$, as can be seen from its representation as the $p$-fold Cartesian product

$$\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p) = \mathcal{H}_5(\mathbb{R}) \times \ldots \times \mathcal{H}_5(\mathbb{R}) = \mathcal{H}_5(\mathbb{R})^p.$$

**Lemma 3.4.2** The representation $B(\mathbb{R}^p) = \sigma(\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p))$ holds.

**Proof.** By Lemma 3.2.1 we already have $B(\mathbb{R}) = \sigma(\mathcal{H}_1(\mathbb{R})) = \sigma(\mathcal{H}_5(\mathbb{R}))$. Using these facts and the representation $\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p) = \mathcal{H}_5(\mathbb{R}) \times \ldots \times \mathcal{H}_5(\mathbb{R})$, we readily get

$$\begin{align*}
\sigma(\mathcal{R}_{\text{SW-Closed}}(\mathbb{R}^p)) &= \sigma(\mathcal{H}_5(\mathbb{R}) \times \ldots \times \mathcal{H}_5(\mathbb{R})) \\
&= \sigma(\sigma(\mathcal{H}_5(\mathbb{R})) \times \ldots \times \sigma(\mathcal{H}_5(\mathbb{R}))) \quad [\text{By Lemma 3.3.1}] \\
&= \sigma(\sigma(\mathcal{H}_0(\mathbb{R})) \times \ldots \times \sigma(\mathcal{H}_0(\mathbb{R}))) \quad [\text{By Lemma 3.2.1}] \\
&= \sigma(\mathcal{H}_0(\mathbb{R}) \times \ldots \times \mathcal{H}_0(\mathbb{R})) \quad [\text{By Lemma 3.3.1}] \\
&= \sigma(\mathcal{R}_{\text{Open}}(\mathbb{R}^p)) \\
(3.15) &= \sigma(B(\mathbb{R}^p)) \quad [\text{By Definition 3.4.2}].
\end{align*}$$
An inspection of the proof of Lemma 3.4.2 leads to the following observation.

**Lemma 3.4.3** The representation

\[ B(\mathbb{R}^p) = \sigma(B(\mathbb{R} \times \ldots \times B(\mathbb{R})) = B(\mathbb{R}) \otimes \ldots \otimes B(\mathbb{R}) \]

holds.

### 3.5 Borel mappings

We specialize the definitions of Section 3.1 to the situation when the domain of the mapping is a measurable space \((S, \mathcal{S})\) and its range space is \(\mathbb{R}^p\) for some positive integer \(p\): Thus, in Definition 3.1.1 we write \((S, \mathcal{S})\) for \((S_a, \mathcal{S}_a)\), \(S_b = \mathbb{R}^p\) and it is understood (unless specified otherwise) that \(S_b = B(\mathbb{R}^p)\).

**Definition 3.5.1** A mapping \(g : S \rightarrow \mathbb{R}^p\) defined on a measurable space \((S, \mathcal{S})\) is a Borel mapping if it is an \((S, B(\mathbb{R}^p))\)-measurable mapping in the sense of Definition 3.1.1, namely that the conditions

\[ g^{-1}(B) \in \mathcal{S}, \quad B \in B(\mathbb{R}^p) \]

all hold.

With \((S, S_e) = (\mathbb{R}^d, B(\mathbb{R}^d))\) for some positive integer \(d\), Fact 3.1.1 takes the following form.

**Fact 3.5.1** If \(g : S \rightarrow \mathbb{R}^p\) and \(h : \mathbb{R}^p \rightarrow \mathbb{R}^q\) are Borel mappings, then the composition mapping \(h \circ g : S \rightarrow \mathbb{R}^q\) is also a Borel mapping.

In this restricted context Lemma 3.1.2 leads to the following fact which is crucial for understanding the importance of probability distributions.

**Lemma 3.5.1** Let \(\mathcal{H}\) denote any collection of subsets of \(\mathbb{R}^p\) which generates the Borel \(\sigma\)-field \(B(\mathbb{R}^p)\), i.e., \(B(\mathbb{R}^p) = \sigma(\mathcal{H})\). The mapping \(g : S \rightarrow \mathbb{R}^p\) is a Borel mapping if and only if the conditions

\[ g^{-1}(E) \in \mathcal{S}, \quad E \in \mathcal{H} \]

all hold.
We close this section by investigating how the measurability of one-dimensional mappings informs the measurability of vector-valued mappings. We begin by noting that any mapping \( g : S \to \mathbb{R}^p \) can also be viewed as a \( p \)-tuple of mappings \( g_1, \ldots, g_p : S \to \mathbb{R} \) where for each \( k = 1, \ldots, p \), the mapping \( g_k : S \to \mathbb{R} \) picks up the \( k \)th coordinate of \( g(s) \) so that
\[
g(s) = (g_1(s), \ldots, g_p(s)), \quad s \in S.
\]

**Lemma 3.5.2** The mapping \( g : S \to \mathbb{R}^p \) is a Borel mapping if and only if the mappings \( g_1, \ldots, g_p : S \to \mathbb{R} \) are all Borel mappings.

**Proof.** We begin with an easy observation: Consider the rectangle \( R \) given by
\[
(3.19) \quad R \equiv B_1 \times B_2 \times \ldots \times B_p
\]
where \( B_1, B_2, \ldots, B_p \) are subsets of \( \mathbb{R} \). It is elementary to see that
\[
g^{-1}(R) = \{ s \in S : g(s) \in B \}
= \{ s \in S : g_{\ell}(s) \in B_{\ell}, \, \ell = 1, \ldots, p \}
= \cap_{\ell=1}^{p} \{ s \in S : g_{\ell}(s) \in B_{\ell} \}
= \cap_{\ell=1}^{p} g_{\ell}^{-1}(B_{\ell}).
\]

First assume that the mapping \( g : S \to \mathbb{R}^p \) is a Borel mapping according to Definition 3.5.1. Fix \( k = 1, \ldots, p \) and use (3.20) with \( B_{\ell} = \mathbb{R} \) for \( \ell = 1, \ldots, p \) whenever \( \ell \neq k \) and \( B_k = (-\infty, a_k) \) for some \( a_k \) in \( \mathbb{R} \). It is plain that a set so constructed is also a Borel subset of \( \mathbb{R}^p \). – See the proof of Lemma 3.4.2 where it is shown (among other things) that \( \sigma(\mathcal{H}_5(\mathbb{R})) \times \ldots \times \sigma(\mathcal{H}_5(\mathbb{R})) \) coincides with \( \mathcal{B}(\mathbb{R}^p) \). It is also plain that \( g^{-1}(R) = g_k^{-1}((-\infty, a_k)) \) since for each \( \ell \neq k \) we have \( g_{\ell}^{-1}(B_{\ell}) = g_{\ell}^{-1}(\mathbb{R}) = S \). But \( g^{-1}(R) \) being an element of \( S \) by the Borel measurability of \( g \), it follows that \( g_k^{-1}((-\infty, a_k)) \) is also an element of \( S \). Therefore, \( a_k \) being arbitrary we conclude that \( g_k^{-1}(\mathcal{H}_0(\mathbb{R})) \subseteq S \), hence \( g_k : S \to \mathbb{R} \) is Borel measurable by virtue of Lemma 3.5.1 and the fact that \( \mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H}_0(\mathbb{R})) \) by Lemma 3.2.1.

Conversely, assume that the mappings \( g_1, \ldots, g_p : S \to \mathbb{R} \) are all Borel mappings. By Lemma 3.4.3, the subset \( R \) defined at (3.19) will be a Borel subset of \( \mathbb{R}^p \) whenever the sets \( B_1, \ldots, B_p \) are all Borel subsets in \( \mathbb{R} \). In that case the assumed Borel measurability of the component mappings implies that the sets \( g_1^{-1}(B_1), \ldots, g_p^{-1}(B_p) \) all belongs to \( S \), whence \( \cap_{\ell=1}^{p} g_{\ell}^{-1}(B_{\ell}) \) also belongs to \( S \) and we conclude that \( g^{-1}(R) \) is an element of \( S \).
3.6 Extended Borel mappings and limits

Sometimes the notion of a Borel mapping defined in Section 3.5 will fail to cover important situations that arise in applications when the mapping can assume the values $\pm \infty$. Recall that the extended real line is defined as $\overline{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{-\infty, +\infty\}$.

**Definition 3.6.1**

Consider a measurable space $(S, \mathcal{S})$. A mapping $g : S \to \overline{\mathbb{R}}$ is said to be an (extended) Borel mapping if

$$g^{-1}(B) \in S, \quad B \in \mathcal{B}(\overline{\mathbb{R}})$$

where the extended Borel $\sigma$-field $\mathcal{B}(\overline{\mathbb{R}})$ on $\overline{\mathbb{R}}$ is simply defined as

$$(3.21) \quad \mathcal{B}(\overline{\mathbb{R}}) \equiv \sigma(\mathcal{B}(\mathbb{R}), \{-\infty\}, \{+\infty\}).$$

It is easy to check that the Borel measurability of the mapping $g : S \to \mathcal{B}(\overline{\mathbb{R}})$ according to Definition 3.6.1 is equivalent to the conditions

$$\{s \in S : g(s) \in (-\infty, a]\} \in \mathcal{S}, \quad a \in \mathbb{R}$$

and

$$S_{\pm \infty} \equiv \{s \in S : g(s) = \pm \infty\} \in \mathcal{S}$$

holding. Furthermore, any Borel mapping $g : S \to \mathbb{R}$ is necessarily an (extended) Borel mapping [Exercise 3.9]

**Lemma 3.6.1** Consider a sequence of extended Borel mappings $\{g_n, n = 1, 2, \ldots\}$ which are all defined on the same measurable space $(S, \mathcal{S})$. The following mappings $S \to [-\infty, \infty]$ derived from the sequence $\{g_n, n = 1, 2, \ldots\}$ are all Borel mappings in the extended sense:

(i) The supremum mapping $S \to \overline{\mathbb{R}}$ defined by

$$s \to \sup_{m \geq 1} g_m(s), \quad s \in S.$$

(ii) The infimum mapping $S \to \overline{\mathbb{R}}$ defined by

$$s \to \inf_{m \geq 1} g_m(s), \quad s \in S.$$
(iii) The limsup mapping \( S \to \mathbb{R} \) defined by
\[
s \mapsto \limsup_{n \to \infty} g_n(s), \quad s \in S.
\]

(iv) The liminf mapping \( S \to \mathbb{R} \) defined by
\[
s \mapsto \liminf_{n \to \infty} g_n(s), \quad s \in S.
\]

**Proof.** For each \( a \) in \( \mathbb{R} \), we note that
\[
\left\{ s \in S : \sup_{m \geq 1} g_m(s) \leq a \right\} = \bigcap_{m \geq 1} \left\{ s \in S : g_m(s) \leq a \right\} \in \mathcal{S}
\]
since for each \( m = 1, 2, \ldots \), the mapping \( g_m : S \to \mathbb{R} \) is an extended Borel mapping with
\[
\{ s \in S : g_m(s) \leq a \} \in \mathcal{S}.
\]
The Borel measurability of the supremum mapping follows from Lemma 3.5.1 as we note that the closed unbounded intervals \( \mathcal{H}_5(\mathbb{R}) = \{(-\infty, a], \ a \in \mathbb{R}\} \) generate the \( \sigma \)-field \( \mathcal{B}(\mathbb{R}) \).

The Borel measurability of the infimum mapping is an immediate consequence of the Borel measurability of the supremum mapping (applied to the mappings \( s \mapsto -g_m(s) \)) upon noting that
\[
\inf_{m \geq 1} g_m(s) = -\sup_{m \geq 1} (-g_m(s)), \quad s \in S.
\]

[Exercise 3.10]. The Borel measurability of the limsup and liminf mappings is now straightforward; the details of the proof are left to the interested reader. \( \blacksquare \)

It is a simple matter to check [Exercise 3.10] that
\[
S^* \equiv \left\{ s \in S : \liminf_{n \to \infty} g_n(s) = \limsup_{n \to \infty} g_n(s) \right\} \in \mathcal{S}
\]
and that on the set \( S^* \), the limit \( \lim_{n \to \infty} g_n(s) \) exists (possibly as an element in \( \mathbb{R} \)), and is the common value assumed by \( \lim \inf_{n \to \infty} g_n \) and \( \lim \sup_{n \to \infty} g_n \).
3.7 Exercises

Ex. 3.1 In the proof of Lemma 3.1.1 show that
a. the collection $g^{-1}(S_b)$ is a $\sigma$-field on $S_a$ if $S_b$ is a $\sigma$-field on $S_b$.

b. the collection $\mathcal{H}_b$ defined at (3.5) is a $\sigma$-field on $S_b$.

Ex. 3.2 Prove the validity of (3.3) and fill in the details of the proof of Fact 3.1.1.

Ex. 3.3 Consider the setting of Definition 3.1.1: Let $\mu_a : S_a \rightarrow [0, \infty]$ a measure defined on $S_a$, and define the set function $\mu_b : S_b \rightarrow [0, \infty]$ by

$$\mu_b(E_b) \equiv \mu_a(g^{-1}(E_b)), \quad E_b \in S_b.$$ 

Show that the set function $\mu_b : S_b \rightarrow [0, \infty]$ is a measure defined on $S_b$. It is a probability measure if $\mu_a$ is a probability measure.

Ex. 3.4 Let $E_a$ and $E_b$ be strict non-empty subsets of the sets $S_a$ and $S_b$, respectively. Show that the complement of $E_a \times E_b$ in $S_a \times S_b$ is usually not a rectangle with sides in $S_a$ and $S_b$, i.e., $E_a \times E_b$ cannot be written as $G_a \times G_b$ with $G_a$ and $G_b$ subsets of $S_a$ and $S_b$.

Use this fact to conclude that if $\mathcal{H}_a$ and $\mathcal{H}_b$ are collections of subsets of $S_a$ and $S_b$, respectively, then the collection $\mathcal{H}_a \times \mathcal{H}_b$ cannot be a field ($\sigma$-field) on $S_a \times S_b$ even if $\mathcal{H}_a$ and $\mathcal{H}_b$ are fields ($\sigma$-fields) on $S_a$ and $S_b$, respectively.

Ex. 3.5 Show that the collections $\mathcal{H}_a^*$ and $\mathcal{H}_b^*$ defined at (3.12) and (3.13), respectively, are $\sigma$-fields.

Ex. 3.6 Give a proof of Lemma 3.4.1.

Ex. 3.7 Give a proof of Lemma 3.4.2.

Ex. 3.8 Most (if not all) mappings $\mathbb{R}^p \rightarrow \mathbb{R}^q$ encountered in applications are Borel mappings. In particular, any continuous mapping $\mathbb{R}^p \rightarrow \mathbb{R}^q$ can be shown to be a Borel mapping! [HINT: Use the fact that a mapping $g : \mathbb{R}^p \rightarrow \mathbb{R}^q$ is continuous if and only if $g^{-1}(\mathcal{O}(\mathbb{R}^q)) \subseteq \mathcal{O}(\mathbb{R}^p)$].

Ex. 3.9 Show the validity of the following statements:

a. The collection $\mathcal{B}(\overline{\mathbb{R}})$ of subsets of $\overline{\mathbb{R}}$ defined by (3.21) is a $\sigma$-field on $\overline{\mathbb{R}}$.

b. Any Borel mapping $g : S \rightarrow \mathbb{R}$ is necessarily an (extended) Borel mapping $S \rightarrow \overline{\mathbb{R}}$. 
Ex. 3.10 Consider the extended Borel mappings $g, h : S \to \overline{\mathbb{R}}$ defined on the same measurable space $(S, \mathcal{S})$.

a. Show that the mapping $S \to \overline{\mathbb{R}} : s \mapsto -g(s)$ is also an extended Borel mapping.

b. Show that the set $S^* = \{s \in S : g(s) = h(s)\}$ belongs to $\mathcal{S}$. 