Part I

PROBABILITY MODELS
Chapter 1

Modeling random experiments

A *random experiment* $E$ is understood as an activity with the following characteristics: It typically has multiple possible outcomes, and the outcome of any realization of the experiment can only be known once the experiment has been realized. Classical examples include the throw of a dice, the fluctuation of the price of a commodity on some stock exchange, etc.

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

- The set $\Omega$ lists all outcomes (samples) generated by the experiment $E$, and is known as the *sample space* for the experiment.

- Events are collections of outcomes, and so are subsets of $\Omega$. The collection of events whose likelihood of occurrence can be defined 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 $\mathcal{P}(\Omega)$.

- The “likelihood” of occurrence of events is assigned only for 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**

With \( \Omega \) an arbitrary set, let \( P(\Omega) \) denote the collection of all subsets of \( \Omega \) (including the empty set) – We often refer to \( P(\Omega) \) as the power set of \( \Omega \) (sometimes also denoted \( 2^{\Omega} \)). Also, let \( \mathcal{F} \) denote a non-empty collection of subsets of \( \Omega \), so \( \mathcal{F} \subseteq P(\Omega) \).

**Definition 1.1.1** The collection \( \mathcal{F} \) is said to be a field (also known as a algebra) on \( \Omega \) if

(F1) \( \emptyset \in \mathcal{F} \)

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

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

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

(F3b) Closed under intersection: If \( E \in \mathcal{F} \) and \( F \in \mathcal{F} \), then \( E \cap F \in \mathcal{F} \)

(F3c) Closed under differences: If \( E \in \mathcal{F} \) and \( F \in \mathcal{F} \), then \( E - F \in \mathcal{F} \) and \( E \Delta F \in \mathcal{F} \)

Note that (F3) implies the seemingly more general statement

(F4) Closed under finite union: If \( E_1, \ldots, E_n \in \mathcal{F} \), then \( \bigcup_{i=1}^{n} E_i \in \mathcal{F} \)

while (F3b) implies the seemingly more general statement

(F4b) Closed under finite intersection: If \( E_1, \ldots, E_n \in \mathcal{F} \), then \( \bigcap_{i=1}^{n} E_i \in \mathcal{F} \)

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

**Definition 1.1.2** The non-empty collection of \( \mathcal{F} \) of subsets of \( \Omega \) is a \( \sigma \)-field (also known as an \( \sigma \)-algebra) on \( \Omega \) if

(F1) \( \emptyset \in \mathcal{F} \)

(F2) Closed under complementarity: If \( E \in \mathcal{F} \), then \( E^c \in \mathcal{F} \)
1.2. PROBABILITY MEASURES

(F3) Closed under countable union: With $I$ a countable index set, if $E_i \in \mathcal{F}$ for each $i \in I$, then $\bigcup_{i \in I} E_i \in \mathcal{F}$

It is plain that any $\sigma$-field is always a field. Any set $\Omega$ always contains two $\sigma$-fields, namely the trivial $\sigma$-field $\{\emptyset, \Omega\}$, and the full $\sigma$-field $\mathcal{P}(\Omega)$.

When $\Omega$ is an arbitrary set and $\mathcal{F}$ is a $\sigma$-field on $\Omega$, it is customary to refer to the pair $(\Omega, \mathcal{F})$ as a measurable space. This is meant to suggest that it is now possible to “measure” the sets in $\mathcal{F}$ by means of measures defined on $\mathcal{F}$.

1.2 Probability measures

Likelihood assignments are implemented through probability measures which we now introduce formally:

Definition 1.2.1 Consider an arbitrary non-empty set $\Omega$ equipped with a $\sigma$-field $\mathcal{F}$. A probability (measure) $\mathbb{P}$ on $\mathcal{F}$ (or on $(\Omega, \mathcal{F})$) is a mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

(P1) $\mathbb{P}[\emptyset] = 0$ and $\mathbb{P}[\Omega] = 1$

(P2) $\sigma$-additivity: With $I$ a countable index set, if $E_i \in \mathcal{F}$ for each $i \in I$, then

$$\mathbb{P}[\bigcup_{i \in I} E_i] = \sum_{i \in I} \mathbb{P}[E_i]$$

whenever the subsets $\{E_i, i \in I\}$ are pairwise disjoint, namely

$$E_i \cap E_j = \emptyset, \quad i \neq j, \quad i, j \in I.$$ 

By (P2) we note that $\mathbb{P}[\Omega] = 1$ implies $\mathbb{P}[\emptyset] = 0$ since $\Omega = \Omega \cup \emptyset$. If (P2) only holds for collections $\{E_i, i \in I\}$ of events with finite $I$, we shall say that the mapping $\mathbb{P} : \mathcal{F} \to [0, 1]$ is (finitely) additive – Obviously, $\sigma$-additivity implies additivity.

Elementary consequences Here are simple consequences of the definitions (F1)-(F5) and (P1)-(P2); proofs are elementary and left to the interested reader as exercises. With events $E$ and $F$ in $\mathcal{F}$, we have
• Complementarity:
\[ P[E^c] = 1 - P[E] \]

• Generalizing additivity:
\[ P[E \cup F] = P[E] + P[F] - P[E \cap F] \]

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

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

Bounds
With countable index set \( I \), let \( \{E_i : i \in I\} \) denote a countable collection of events in \( \mathcal{F} \). The following elementary bounds can be established by induction:

• Boole’s inequality (also known as union bound) is commonly used in Information Theory and theoretical Computer Science, and states that
\[ P[\cup_{i \in I} E_i] \leq \sum_{i \in I} P[E_i]. \]

• Bonferroni’s inequality gives a lower bound to \( P[\cup_{i \in I} E_i] \): With finite index set \( I \), it holds that
\[ P[\cup_{i \in I} E_i] \geq \sum_{i \in I} P[E_i] - \sum_{i,j \in I : i < j} P[E_i \cap E_j]. \]

Both bounds are shown by induction on the size \( |I| \) of the collection. The union bound is first established when \( I \) is finite; the countably infinite case then follows by an application of Lemma 1.4.1.

In many applications a major question is concerned with generating the probability measure \( P \) that captures the salient features of the experiment \( \mathcal{E} \) under consideration once its sample space \( \Omega \) has been identified. In particular, this requires that the \( \sigma \)-field \( \mathcal{F} \) of events be judiciously chosen. There are a number of ways to do so, and we discuss one approach in the next section.
1.3 Discrete probability models

A case of particular importance arises when $\Omega$ is countable, in which case it is customary to take $\mathcal{F} = \mathcal{P}(\Omega)$. In that setting, specifying $\mathbb{P}$ on $(\Omega, \mathcal{P}(\Omega))$ is equivalent to specifying

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

This is a straightforward consequence of the $\sigma$-additivity of probability measures and the fact that

$$F = \bigcup_{\omega \in F} \{\omega\}, \quad F \in \mathcal{P}(\Omega).$$ 

Indeed,

$$\mathbb{P}[F] = \sum_{\omega \in F} p(\omega)$$

as we set

$$p(\omega) \equiv \mathbb{P}[[\omega]], \quad \omega \in \Omega.$$

**Uniform probability assignments** Let $\Omega$ be an arbitrary set to be used as the sample space of a probabilistic experiment $\mathcal{E}$ where outcomes are equally likely to occur – Thus, according to $\mathcal{E}$ an element of $\Omega$ is selected at random as the saying goes, or more accurately, uniformly. A natural question is how to define the corresponding probability measure $\mathbb{P}$, hereafter referred to as the uniform probability measure. We consider several cases.

(i) The set $\Omega$ is a discrete set with a finite number of elements, say $\Omega = \{\omega_1, \ldots, \omega_N\}$ for some finite $N$. The uniform probability measure on such set $\Omega$ assigns the same probability of occurrence to any outcome. Thus, $p(\omega_1) = \ldots = p(\omega_N) = p$, so that

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

whence

$$p = \frac{1}{|\Omega|}$$

upon taking $F = \Omega$. Finally we get

$$\mathbb{P}[F] = \frac{|F|}{|\Omega|}, \quad F \in \mathcal{P}(\Omega).$$
(ii) The set $\Omega$ is discrete with $|\Omega| = \infty$, say $\Omega = \{\omega_n, n = 1, 2, \ldots\}$: We should still take $p(\omega)$ to be independent of $\omega$, say $p(\omega) = p$ for all $\omega \in \Omega$ where $p$ is this common value. It still follows that

$$P[F] = |F|p, \quad F \in \mathcal{P}(\Omega)$$

Therefore, we get

$$|F|p \leq 1, \quad F \in \mathcal{P}(\Omega)$$

and this implies $p = 0$ (because we can select a sequence $\{F_n, n = 1, 2, \ldots\}$ of subsets of $\Omega$ such that $|F_n| = n$ for all $n = 1, 2, \ldots$ – Just take $F_n = \{\omega_1, \ldots, \omega_n\}$). A contradiction immediately arises since by $\sigma$-additivity we have

$$1 = P[\Omega] = \sum_{\omega \in \Omega} p = 0!$$

In other words, it is not possible to have a uniform probability measure on a discrete set with $|\Omega| = \infty$.

What happens when $\Omega$ is uncountable? In Chapter 2 we shall see that for the purpose of defining probability measures on non-countable sets $\Omega$, in general it is not possible to take $\mathcal{F} = \mathcal{P}(\Omega)$. This is due to the fact that the $\sigma$-additivity of $P$ imposes too many constraints, forcing a reduction of $\mathcal{P}(\Omega)$. In particular, in the non-countable case, it is not possible to assign a likelihood of occurrence (through a probability measure satisfying the axioms (P1)-(P2)) to every subset of $\Omega$! The difficulties involved will be illustrated on two examples, namely infinitely many coin tosses of a fair coin in Section 2.3 and selecting a point at random in the interval $[0, 1]$ in Section 2.4.

### 1.4 Continuity properties of probability measures

Consider a sequence $\{E_n, n = 1, 2, \ldots\}$ of events in $\mathcal{F}$.

The impact of monotonicity
1.4. CONTINUITY PROPERTIES OF PROBABILITY MEASURES

**Lemma 1.4.1** If the sequence is monotone increasing in the sense that

\[ E_n \subseteq E_{n+1}, \quad n = 1, 2, \ldots \]

then \( \lim_{n \to \infty} P[E_n] = P[\bigcup_{n=1}^\infty E_n] \).

**Proof.** Note the relation

\[ \bigcup_{n=1}^\infty E_n = \bigcup_{m=1}^\infty F_m \]

where

\[ F_m \equiv E_m - E_{m-1}, \quad m = 1, 2, \ldots \]

(under the convention \( E_0 = \emptyset \)). The events \( \{F_m, \ m = 1, 2, \ldots\} \) being pairwise disjoint, we get

\[
P[\bigcup_{n=1}^\infty E_n] = \sum_{m=1}^\infty P[F_m] \quad \text{[By the } \sigma\text{-additivity of } P]\]

\[
= \sum_{m=1}^\infty (P[E_m] - P[E_{m-1}])
\]

\[
= \lim_{m \to \infty} \left( \sum_{k=1}^m (P[E_k] - P[E_{k-1}]) \right)
\]

\[
= \lim_{m \to \infty} (P[E_m] - P[E_0]) = \lim_{m \to \infty} P[E_m].
\]

(1.1)

This result can be interpreted as a continuity result for \( P \) in the following sense: If we define \( \lim_{n \to \infty} E_n \equiv \bigcup_{n=1}^\infty E_n \), then Lemma 1.4.1 states that \( \lim_{n \to \infty} P[E_n] = P[\lim_{n \to \infty} E_n] \).

**Lemma 1.4.2** If the sequence is monotone decreasing in the sense that

\[ E_{n+1} \subseteq E_n, \quad n = 1, 2, \ldots \]

then \( \lim_{n \to \infty} P[E_n] = P[\bigcap_{n=1}^\infty E_n] \).

This result can also be recast as a continuity result for \( P \): This time, if we define \( \lim_{n \to \infty} E_n \equiv \bigcap_{n=1}^\infty E_n \), then \( \lim_{n \to \infty} P[E_n] = P[\lim_{n \to \infty} E_n] \) by virtue of
Lemma 1.4.2. The proof of this result is similar to the one given for the case of a monotone increasing sequence of events. In fact, these two results are equivalent once we observe that a sequence \( \{E_n, n = 1, 2, \ldots\} \) is monotone increasing (resp. decreasing) if and only if its complementary sequence \( \{E_n^c, n = 1, 2, \ldots\} \) is monotone decreasing (resp. increasing).

Limsup and liminf, and limits  In analogy with the convergence of sequences on \( \mathbb{R} \), these continuity results for monotone sequences of events can be generalized as follows. Let \( \{E_n, n = 1, 2, \ldots\} \) be a collection of events in \( \mathcal{F} \). Define

\[
\limsup_{n \to \infty} E_n \equiv \bigcap_{n=1}^{\infty} \left( \bigcup_{m=n}^{\infty} E_m \right) = \bigcap_{n=1}^{\infty} \bar{E}_n
\]

with

\[
\bar{E}_n = \bigcup_{m=n}^{\infty} E_m, \quad n = 1, 2, \ldots
\]

Similarly,

\[
\liminf_{n \to \infty} E_n \equiv \bigcup_{n=1}^{\infty} \left( \bigcap_{m=n}^{\infty} E_m \right) = \bigcup_{n=1}^{\infty} E_n
\]

with

\[
E_n = \bigcap_{m=n}^{\infty} E_m, \quad n = 1, 2, \ldots
\]

The events \( \limsup_{n \to \infty} E_n \) and \( \liminf_{n \to \infty} E_n \) always exist, and as expected, we refer to them as the limit sup and limit inf of the collection \( \{E_n, n = 1, 2, \ldots\} \), respectively. We have the mnemonic notation

\[
\limsup_{n \to \infty} E_n = [E_n \text{ infinitely often (i.o.) }]
\]

and

\[
\liminf_{n \to \infty} E_n = [\text{ Eventually all } E_n].
\]

**Definition 1.4.1** The collection \( \{E_n, n = 1, 2, \ldots\} \) of events will be said to converge if the condition

\[
\limsup_{n \to \infty} E_n = \liminf_{n \to \infty} E_n
\]

holds, in which case we shall write

\[
\lim_{n \to \infty} E_n \equiv \limsup_{n \to \infty} E_n = \liminf_{n \to \infty} E_n.
\]
1.4. CONTINUITY PROPERTIES OF PROBABILITY MEASURES

Obviously, for each $n = 1, 2, \ldots$ the inclusion $E_n \subseteq \bar{E}_n$ holds and we have the monotone properties

$$E_{n+1} \subseteq \bar{E}_n \quad \text{and} \quad E_n \subseteq \bar{E}_{n+1}.$$ 

By the continuity of $\mathbb{P}$ on monotone sequences discussed earlier it follows that

$$\mathbb{P}\left[\limsup_{n \to \infty} E_n\right] = \lim_{n \to \infty} \mathbb{P}\left[\bar{E}_n\right]$$

and

$$\mathbb{P}\left[\liminf_{n \to \infty} E_n\right] = \lim_{n \to \infty} \mathbb{P}\left[E_n\right].$$

With the definition of set continuity given above we have the following continuity property for probability measures.

**Lemma 1.4.3** If the collection $\{E_n, n = 1, 2, \ldots\}$ of events in $\mathcal{F}$ converges, then

$$\lim_{n \to \infty} \mathbb{P}[E_n] = \mathbb{P}\left[\lim_{n \to \infty} E_n\right]$$

without any monotonicity assumption on the collection $\{E_n, n = 1, 2, \ldots\}$.

**Proof.** By the remarks above, the convergence of the collection $\{E_n, n = 1, 2, \ldots\}$ implies

$$\mathbb{P}\left[\lim_{n \to \infty} E_n\right] = \lim_{n \to \infty} \mathbb{P}\left[\bar{E}_n\right] \quad \text{and} \quad \mathbb{P}\left[\lim_{n \to \infty} E_n\right] = \lim_{n \to \infty} \mathbb{P}\left[E_n\right].$$

On the other hand, we have

$$E_n \subseteq E_n \subseteq \bar{E}_n, \quad n = 1, 2, \ldots$$

whence

$$\mathbb{P}[E_n] \leq \mathbb{P}[E_n] \leq \mathbb{P}[\bar{E}_n], \quad n = 1, 2, \ldots$$

A standard sandwich argument yields the desired result as we let $n$ go to infinity in this chain of inequalities. ■
1.5 Independence

Consider a collection \( \{E_i, \ i \in I\} \) of events in \( \mathcal{F} \) where \( I \) is an arbitrary index set.

- Pairwise independence: The events \( \{E_i, \ i \in I\} \) are said to be pairwise independent if the conditions
  \[
  \mathbb{P}[E_i \cap E_j] = \mathbb{P}[E_i] \mathbb{P}[E_j], \quad i \neq j
  \]
  hold. When \( I \) is finite, this is a set of \( \frac{|I||I|-1}{2} \) conditions.

- Mutual independence (with \( I \) finite): The events \( \{E_i, \ i \in I\} \) are said to be mutually independent if
  \[
  \mathbb{P}[\bigcap_{j \in J} E_j] = \prod_{j \in J} \mathbb{P}[E_j], \quad |J| > 0.
  \]
  This represents \( 2^{|I|} - (|I| + 1) \) non-trivial conditions.

- Mutual independence (with \( I \) arbitrary): The events \( \{E_i, \ i \in I\} \) are said to be mutually independent if for each finite subset \( J \subseteq I \) with \( 0 < |J| < \infty \), the events \( \{E_j, \ j \in J\} \) are mutually independent, namely
  \[
  \mathbb{P}[\bigcap_{j \in J} E_j] = \prod_{j \in J} \mathbb{P}[E_j], \quad 0 < |J| < \infty.
  \]

Set-theoretic operations preserve independence in the following sense.

**Theorem 1.5.1** 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) The events \( \{E_i^*, \ i \in I\} \) are mutually independent where for each \( i \in I \), \( E_i^* \) is either \( E_i \) or its complement \( E_i^c \).

(iii) 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 take the complement of a set, union and intersection.
1.6. BOREL-CANTELLI LEMMAS

Part (i) is trivial, and Part (ii) is subsumed by Part (iii); the proof of this fact is rather cumbersome and will not be given here. With (i) in mind we note that pairwise independence does not imply mutual independence; this is already apparent from the following simple example.

Example 1.5.1 Consider the experiment where an item is selected uniformly from a set containing four distinct objects labelled $a, b, c, d$. Thus, $\Omega = \{a, b, c, d\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}$ is fully determined by the assignment $p(a) = p(b) = p(c) = p(d) = \frac{1}{4}$. Now define, the events $A = \{a, b\}$, $B = \{b, c\}$ and $C = \{a, c\}$ so that $\mathbb{P}[A] = \mathbb{P}[B] = \mathbb{P}[C] = \frac{1}{2}$. It is plain that $\mathbb{P}[A \cap B] = \mathbb{P}[B \cap C] = \mathbb{P}[A \cap C] = \frac{1}{4}$, and the events $A$, $B$ and $C$ are indeed pairwise independent. However, $A \cap B \cap C = \emptyset$, whence $\mathbb{P}[A \cap B \cap C] = 0 \neq \frac{1}{8}$ and the $A$, $B$ and $C$ are not mutually independent.

1.6 Borel-Cantelli Lemmas

The Borel-Cantelli lemmas given next provide an example of a zero-one law. Recall that if $\{A_n, n = 1, 2, \ldots\}$ is a collection of events in $\mathcal{F}$, then

$$\limsup_{n \to \infty} A_n = [A_n \text{ i.o.}] = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

Lemma 1.6.1 If $\{A_n, n = 1, 2, \ldots\}$ is a collection of events in $\mathcal{F}$ such that

$$\sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty,$$

then it is always the case that $\mathbb{P}[A_n \text{ i.o.}] = 0$.

Proof. Obviously,

$$\mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}\left[\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right]$$

$$= \lim_{n \to \infty} \mathbb{P}\left[\bigcup_{m=n}^{\infty} A_m\right] \quad \text{[By monotonicity in $n$]}$$

$$= \lim_{n \to \infty} \left( \lim_{k \to \infty} \mathbb{P}\left[\bigcup_{m=n}^{k} A_m\right] \right) \quad \text{[By monotonicity in $k$]}$$

$$\leq \lim_{n \to \infty} \left( \lim_{k \to \infty} \sum_{m=n}^{k} \mathbb{P}[A_m] \right) \quad \text{[By union bound on $\mathbb{P}\left[\bigcup_{m=n}^{k} A_m\right]$]}$$
\[
\lim_{n \to \infty} \left( \sum_{m=n}^{\infty} \mathbb{P}[A_m] \right).
\]

The result follows since

\[
\lim_{n \to \infty} \left( \sum_{m=n}^{\infty} \mathbb{P}[A_m] \right) = 0
\]

under the convergence condition \( \sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty \).

It is natural to wonder what happens to the conclusion of Lemma 1.6.2 when the condition \( \sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty \) holds instead. If we add independence, then the following result holds.

**Lemma 1.6.2** When the events \( \{A_n, \ n = 1, 2, \ldots\} \) in \( \mathcal{F} \) are mutually independent, then \( \mathbb{P}[A_n \text{ i.o.}] = 1 \) under the condition

\[
\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty.
\]

**Proof.** Our point of departure is the observation that

\[ [A_n \text{ i.o.}]^c = \bigcup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c. \]

By arguments similar to the one given in the proof of Lemma 1.6.1

\[
1 - \mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}[\bigcup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c]
= \lim_{n \to \infty} \mathbb{P}[\cap_{m=n}^{\infty} A_m^c] \quad \text{[By monotonicity in } n]\]
= \lim_{n \to \infty} \left( \lim_{k \to \infty} \mathbb{P}[\cap_{m=n}^{k} A_m^c] \right) \quad \text{[By monotonicity in } k].
\]

For each \( n = 1, 2, \ldots \) and \( k = n, n+1, \ldots \), we see that

\[
\mathbb{P}[\cap_{m=n}^{k} A_m^c] = \prod_{m=n}^{k} \mathbb{P}[A_m^c] \quad \text{[By independence]}
\]
1.7. CONDITIONAL PROBABILITIES

\[ = \prod_{m=n}^{k} (1 - \mathbb{P}[A_m]) \]

\[ = \prod_{m=n}^{k} e^{-\mathbb{P}[A_m]} \quad \text{[Because } 1 - x \leq e^{-x}, \, x \geq 0 \text{]} \]

\[ = e^{-\sum_{m=n}^{k} \mathbb{P}[A_m]} . \]

(1.3)

Thus, \( \lim_{k \to \infty} \mathbb{P}[\bigcap_{m=n}^{k} A_m^c] = 0 \) since \( \sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty \), and the desired conclusion follows.

Without the assumption of independence the conclusion of Lemma 1.6.2 may not hold as the following example shows.

Example 1.6.1

1.7 Conditional probabilities

Conditional probabilities are what is needed when there is no independence. We begin with a classical definition: With \( A \) and \( B \) events in \( \mathcal{F} \) such that \( \mathbb{P}[B] > 0 \), define the \textit{conditional probability of } \( A \text{ given } B \) by

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

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

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

is always true regardless of \( \mathbb{P}[B] > 0 \) or not.

When \( \mathbb{P}[B] > 0 \) we can define the mapping \( Q_B : \mathcal{F} \to [0, 1] \) by

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

It is easy to show that \( Q_B : \mathcal{F} \to \mathbb{R}_+ \) 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 \( \mathbb{P}[B] = 0 \).
CHAPTER 1. MODELING RANDOM EXPERIMENTS

Here are three easy consequences: With $I$ a countable index set, let $\{B_i, \ i \in I\}$ be events in $\mathcal{F}$ that form a partition of $\Omega$, i.e.,

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

The law of total probabilities

Because $A = \cup_{i \in I} (A \cap B_i)$, we have

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

(1.4)

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

Put differently,

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

Bayes’ rule – From prior probabilities to posterior probabilities

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

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

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

(1.5)

$$= \frac{\sum_{i \in I} P[A \cap B_i]}{\sum_{i \in I} P[A|B_i] P[B_i]}.$$  

Modeling sequential decision making

If $I$ is a finite set, say $I = \{1, \ldots, n\}$, we have

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

This can shown by induction on $n$. 
Chapter 2

Measurability

As already discussed in Chapter 1, determining the appropriate $\sigma$-field $F$ of events is technically more delicate when $\Omega$ is countably infinite. This is in large part due to measure-theoretic constraints, i.e., constraints that are imposed by the $\sigma$-additivity of probability measures to be defined on $F$. These issues are briefly explored in the present chapter.

2.1 Generating $\sigma$-fields

In order to formalize these ideas, introduce several useful definitions:

Fact 2.1.1 Let $\{S_i, i \in I\}$ be a non-empty family of $\sigma$-fields on some set $S$ with $I$ an arbitrary index set. The collection of subsets of $S$ defined by

$$\{E \in \mathcal{P}(S) : E \in S_i, i \in I\}$$

is a $\sigma$-field on $S$ denoted indifferently either $\bigwedge_{i \in I} S_i$ or $\bigwedge (S_i, i \in I)$. We refer to it as the intersection of the $\sigma$-fields $\{S_i, i \in I\}$.

The proof that $\bigwedge_{i \in I} F_i$ is a $\sigma$-field on $S$ is left as an exercise.

Definition 2.1.1 If $\mathcal{G}$ is a collection of subsets of $S$, let $\sigma(\mathcal{G})$ denote the smallest $\sigma$-field on $S$ that contains $\mathcal{G}$.

This definition is well posed as we note that

$$\sigma(\mathcal{G}) = \bigwedge \left( S \subseteq \mathcal{P}(S) : \begin{array}{c} S \text{ is a } \sigma\text{-field on } S \\ \text{and} \\ \mathcal{G} \text{ is contained in } S \end{array} \right).$$
CHAPTER 2. MEASURABILITY

This collection is not empty because $\mathcal{P}(S)$ is a $\sigma$-field that contains $\mathcal{G}$.

**Definition 2.1.2** If $\mathcal{G}$ and $\mathcal{S}$ are two collections of subsets of $S$ with $\mathcal{G} \subseteq \mathcal{S}$, whenever $\mathcal{S}$ is a $\sigma$-field on $S$, we say that $\mathcal{G}$ generates the $\sigma$-field $\mathcal{S}$, or equivalently, that $\mathcal{G}$ is a generating family (or a generator) for $\mathcal{S}$, whenever $\mathcal{S} = \sigma(\mathcal{G})$.

The following fact is elementary.

**Fact 2.1.2** If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two collections of subsets of $S$ such that $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $\sigma(\mathcal{G}_1) \subseteq \sigma(\mathcal{G}_2)$.

If $A$ is a subset of $S$ and $\mathcal{G}$ is a collection of subsets of $S$, then the *trace* $\mathcal{G} \cap A$ of $\mathcal{G}$ on $A$ is the collection of subsets of $A$ given by

$$\mathcal{G} \cap A \equiv \{ A \cap G : G \in \mathcal{G} \}.$$  

**Fact 2.1.3** The collection $\mathcal{G} \cap A$ is a $\sigma$-field on $A$ if $\mathcal{G}$ is a $\sigma$-field on $S$. Moreover, if $A$ is itself a member of $\mathcal{G}$, then $\mathcal{G} \cap A$ is also given by

$$\mathcal{G} \cap A = \{ G \in \mathcal{G} : G \subseteq A \}.$$  

Next, consider a mapping $g : S_a \rightarrow S_b$ where $S_a$ and $S_b$ are two arbitrary sets (possibly identical). Let $\mathcal{H}$ be a collection of subsets of $S_b$ (so $\mathcal{H} \subseteq \mathcal{P}(S_b)$). Define

$$g^{-1}(\mathcal{H}) = \{ g^{-1}(H_b) : H_b \in \mathcal{H} \}$$

where as usual we have

$$g^{-1}(H_b) = \{ x \in S_a : g(x) \in H_b \}, \quad H_b \in \mathcal{P}(S_b).$$

**Lemma 2.1.1** The collection $g^{-1}(\mathcal{H})$ of subsets of $S_a$ is a $\sigma$-field on $S_a$ if $\mathcal{H}$ is a $\sigma$-field on $S_b$. More generally, for any collection $\mathcal{H}$ of subsets of $S_b$, we have

$$g^{-1}(\sigma(\mathcal{H})) = \sigma(g^{-1}(\mathcal{H})).$$

**Proof.** The inclusion $\sigma(g^{-1}(\mathcal{H})) \subseteq g^{-1}(\sigma(\mathcal{H}))$ is straightforward since $g^{-1}(\sigma(\mathcal{H}))$ is a $\sigma$-field which contains $g^{-1}(\mathcal{H})$. To establish the reverse inclusion $g^{-1}(\sigma(\mathcal{H})) \subseteq \sigma(g^{-1}(\mathcal{H}))$, consider the collection

$$\mathcal{H}_g \equiv \{ H_b \in \mathcal{P}(S_b) : g^{-1}(H_b) \in \sigma(g^{-1}(\mathcal{H})) \}.$$  

It is easy to check that $\mathcal{H}_g$ is a $\sigma$-field on $S_b$ and that it contains $\mathcal{H}$. Therefore, $\mathcal{H}_g$ contains $\sigma(\mathcal{H})$, and the conclusion $g^{-1}(\sigma(\mathcal{H})) \subseteq \sigma(g^{-1}(\mathcal{H}))$ follows.  

□
2.2 Extensions of probability measures

It is often the case that when identifying $\mathcal{F}$ and $\mathbb{P}$, there is a collection $\mathcal{G}$ of subsets of $\Omega$ where $\mathbb{P}$ is naturally defined. In such situations it is then natural to consider the smaller $\sigma$-field that contains $\mathcal{G}$ with the hope that the probability measure can be uniquely extended to it.

In constructing a probability model $(\Omega, \mathcal{F}, \mathbb{P})$ for a random experiment $\mathcal{E}$ we are often faced with the following situation: After identifying the sample space $\Omega$, structural properties of $\mathcal{E}$ naturally suggest likelihood assignments for events in a collection $\mathcal{G}$ – Let $\text{Lik}(G)$ denote the likelihood of event $G$ in $\mathcal{G}$. Two points should be clear:

(i) We expect that the desired $\sigma$-field on $\Omega$ would contain $\mathcal{G}$ – In fact, in the name of minimality it would be natural to require $\mathcal{F} \equiv \sigma(\mathcal{G})$. After all, $\sigma(\mathcal{G})$ is the smallest $\sigma$-field where a probability measure $\mathbb{P}$ could be defined that is compatible with the probability assignments $\{\text{Lik}(G), G \in \mathcal{G}\}$.

(ii) The probability assignments $\{\text{Lik}(G), G \in \mathcal{G}\}$ being the values taken by the probability measure to be defined on the events in $\mathcal{G}$ (and then ultimately, on $\sigma(\mathcal{G})$), it stands to reason that we should define $\mathbb{P}$ on $\mathcal{G}$ by

(2.1) \[ \mathbb{P}[G] \equiv \text{Lik}(G), \quad G \in \mathcal{G}. \]

However, such a definition should be compatible with the axioms (F1)-(F2) satisfied by probability measures. In particular, with $I$ a countable index set, if the sets $\{G_i, i \in I\}$ are pairwise disjoint with $G_i \in \mathcal{G}$ for each $i \in I$, then one of two possibilities can occur:

(a) If $\bigcup_{i \in I} G_i$ is an element of $\mathcal{G}$ as well, then the relation

(2.2) \[ \text{Lik}(\bigcup_{i \in I} G_i) = \sum_{i \in I} \text{Lik}(G_i) \]

must hold since (2.1) implies

(2.3) \[ \mathbb{P}[\bigcup_{i \in I} G_i] = \sum_{i \in I} \mathbb{P}[G_i] \]

as required by the $\sigma$-additivity of probability measures. Similarly, if both $G$ and $G^c$ are in $\mathcal{G}$, it should be the case that

(2.4) \[ \text{Lik}(G^c) = 1 - \text{Lik}(G), \]
to reflect the complementarity property
\[ P[G^c] = 1 - P[G] \]  
by virtue of (2.1).

(b) On the other hand, if \( \bigcup_{i \in I} G_i \) is not an element of \( G \), then it is natural to define its likelihood of such an event by setting
\[ P[\bigcup_{i \in I} G_i] = \sum_{i \in I} \text{Lik}(G_i). \]  
Similarly, if \( G \) is an element of \( G \) but its complement \( G^c \) is not, then we should set
\[ P[G^c] = 1 - \text{Lik}(G). \]  

In principle this provides a constructive approach to recursively building a probability measure \( P \) on \( \sigma(G) \) by taking further unions and complements of sets to which a likelihood value has been assigned under the probability measures \( P \) as it is being constructed. With this in mind it is natural to wonder whether ultimately a probability measure \( P \) can be constructed on \( \sigma(G) \) which is consistent with the likelihood assignments \( \{ \text{Lik}(G), G \in G \} \). A moment of reflection should convince the reader that conditions are needed on \( G \) for this to occur.

Theorem 2.2.1

2.3 Example 1: Infinite coin tosses

The experiment \( E \) consists in repeating a coin toss under "identical and independent conditions" with a fair coin (so that the likelihood of occurrence of Head is the same as that of Tail). It is convenient to take the sample space \( \Omega \) to be \( \{0, 1\}^{\mathbb{N}_0} \), i.e.,
\[ \Omega = \{ \omega = (\omega_1, \omega_2, \ldots) : \omega_k \in \{0, 1\}, k = 1, 2, \ldots \} \]  
with the understanding that \( \omega_k = 1 \) (resp. \( \omega_k = 0 \)) if the \( k^{th} \) toss yields Head (resp. Tail). Note that \( \Omega \) has the same cardinality as the unit interval \([0, 1]\) (hence is uncountable). How should we construct \( \mathcal{F} \) (and \( P \))?

It is natural to require that for any \( n = 1, 2, \ldots \), any collection of outcomes determined by the first \( n \) tosses should be an event in \( \mathcal{F} \) – After all one should expect that the model we are seeking to construct would also contain a model for
each of the finite toss experiments. In particular, with any given binary sequence \((b_1, \ldots, b_n)\) of length \(n\), consider

\[
F_n(b_1, \ldots, b_n) \equiv \left\{ \omega = (\omega_1, \omega_2, \ldots) \in \Omega : \omega_k = b_k \quad k = 1, \ldots, n \right\}.
\]

It is plain that \(\mathcal{F}\) must at least contain these events which are determined by a finite number of coin tosses, namely

\[
F_n(b_1, \ldots, b_n) \in \mathcal{F}.
\]

Fairness (which is essentially a uniformity condition) requires that

\[
P[F_n(b_1, \ldots, b_n)] = 2^{-n}
\]

since \(P[F_n(b_1, \ldots, b_n)]\) should not depend on \((b_1, \ldots, b_n)\) and there are \(2^n\) distinct sets of the form (2.8). Note also that

\[
\bigcup_{(b_1, \ldots, b_n) \in \{0, 1\}^n} F_n(b_1, \ldots, b_n) = \Omega
\]

with the collection

\[
\mathcal{G} \equiv \left\{ F_n(b_1, \ldots, b_n), \quad b_1, \ldots, b_n \in \{0, 1\}, \quad n = 1, 2, \ldots \right\}
\]

forming a collection of non-overlapping events.

It is therefore natural to take \(\mathcal{F} = \sigma (\mathcal{G})\) where the generator \(\mathcal{G}\) is the collection \(\mathcal{G}\). Although the \(\sigma\)-field \(\mathcal{F}\) so defined is very large, it does not coincide with \(\mathcal{P}(\Omega)\).

It does however contain interesting events that do not depend on a given finite number of tosses. For instance, consider the event \(F\) given by

\[
F = \left\{ \omega = (\omega_1, \omega_2, \ldots) \in \Omega : \text{A even number of tosses is needed before observing the first Head} \right\}
\]

\[
= \bigcup_{k=1}^{\infty} E_{2k}
\]

where for each \(k = 1, \ldots\) we have defined \(E_k = F_k(0, \ldots, 0, 1)\). The event \(F\) is clearly an element of \(\mathcal{F}\).

Measure Theory tells us that that there exists a \emph{unique} probability measure \(\mathbb{P}\) on \(\mathcal{F}\) so that (2.10) holds for all \(n = 1, 2, \ldots\). In particular, it is easy to check that

\[
P[F] = P[\bigcup_{k=1}^{\infty} E_{2k}]
\]
\[ = \sum_{k=1}^{\infty} \mathbb{P}[E_{2k}] \quad \text{[By required } \sigma\text{-additivity]} \]

\[ = \sum_{k=1}^{\infty} 2^{-2k} \]

\[ = \frac{2^{-2}}{1 - 2^{-2}} = \frac{1}{3}. \]

(2.11)

2.4 Example 2: Selecting a point at random in the interval \([0, 1]\)

A particularly important case is that of turning the non-countable interval \([0, 1]\) into a measurable space on which likelihood of occurrence can be defined through a probability measure. Consider the random experiment where a point is selected at random in the finite interval \([0, 1]\), so here it is appropriate to take \(\Omega = [0, 1]\). Intuitively, following the general approach outlined in Section 2.2 we could proceed as follows to define \(\mathcal{F}\) and \(\mathbb{P}\) (denoted here \(\lambda\) for Lebesgue measure).

We begin with a well-known fact of topology on \(\mathbb{R}\): A subset \(U\) of \(\mathbb{R}\) is said to be open if for every \(x\) in \(U\), there exists an interval \(I_x\) of the form \((a_x, b_x)\) such that \(x \in I_x\) and \(I_x \subseteq U\). A set \(F\) is said to be closed if its complement \(F^c\) is open. It is elementary to check that intervals of the form \((a, b)\) (with \(a < b\) in \(\mathbb{R} \cap \{\pm \infty\}\)) are indeed open.

**Fact 2.4.1** 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

\[ U = \bigcup_{i \in I} J_i \quad \text{with} \quad J_k \cap J_\ell = \emptyset, \quad k \neq \ell \quad k, \ell \in I. \]

(2.12)

To define the appropriate \(\sigma\)-field \(\mathcal{F}\) and the probability measure \(\lambda\) on it, it is natural to proceed as follows:

(i) First it is natural to require that singletons be events, i.e., that \(\{\omega\}\) be an element of the \(\sigma\)-field \(\mathcal{F}\) for every \(\omega\) in \(\Omega\). Indeed, the model should allow one to answer questions such as “what is the probability that \(\frac{2}{3}\) was selected?” As in Section 1.3, the assumption of uniform selection again would be recast as requiring
2.4. EXAMPLE 2: SELECTING A POINT AT RANDOM IN THE INTERVAL \([0, 1]\)

\[ \lambda(\{\omega\}) \text{ to be independent of } \omega, \text{ and the arguments developed there also imply} \]

\[ \lambda(\{\omega\}) = 0, \quad \omega \in \Omega \]  

(2.13)

upon using the fact that \(\Omega\) is not countable. It immediately follows that \(\lambda(E) = \) for every countable subset \(E\) of \(\Omega\).

(ii) Thus, in order to extend the definition of \(\lambda\) to non-countable sets, it appears that additional constraints associated with uniform selection need to be leveraged. For instance, if a point is selected uniformly at random in \((0, 1)\), it is natural to assume that the likelihood of selecting a point in an interval \([a, b] \subseteq \Omega\) should depend only on the size of the interval, and not on its location, say

\[ \lambda([a, b]) = b - a. \]  

(2.14)

(iii) Note that (2.13) is compatible with (2.14): We have \(\lambda(\{a\}) = 0\) for all \(a\) in \(\Omega\) by taking \(a = b\) in (2.14). It follows that

\[ \lambda((a, b)) = b - a. \]  

(2.15)

(iv) The union of countable collections of open intervals should be in the \(\sigma\)-field \(\mathcal{F}\). Therefore, by Fact 2.4.1, we see that \(\mathcal{F}\) should include every open set \(U \subseteq [0, 1]\) and \(\sigma\)-additivity requires that

\[ \lambda(U) = \sum_{i \in J} \lambda(J_i) \]

where the notation and the assumptions are the ones used in Fact 2.4.1.

(v) A set \(F\) of \([0, 1]\) being closed if and only \(F^c\) is open, we conclude that every closed set \(F \subseteq (0, 1)\) must also also belong to \(\mathcal{F}\) with \(\lambda(F) = 1 - \lambda(F^c)\).

(vi) Clearly, any countable union of open subsets should be in \(\mathcal{F}\), and any countable intersection of closed subsets should be in \(\mathcal{F}\).

This leads to defining \(\mathcal{F}\) as

\[ \mathcal{F} = \sigma(\mathcal{I}([0, 1])) \]

where \(\mathcal{I}([0, 1])\) denotes the collection of all open intervals contained in \([0, 1]\), with

\[ \mathcal{I}([0, 1]) \equiv \{(a, b) : 0 \leq a < b \leq 1\}. \]

**Definition 2.4.1**  The \(\sigma\)-field \(\sigma(\mathcal{I}([0, 1]))\) is called the Borel \(\sigma\)-field on \([0, 1]\) and is denoted by \(B([0, 1])\).
By Fact 2.4.1 we have also the characterization

\[ \mathcal{B}([0, 1]) = \sigma(\mathcal{O}([0, 1])) \]

where \( \mathcal{O}([0, 1]) \) denotes the collection of all open sets contained in \([0, 1]\). More generally, with \( I \) denoting an interval (closed or open or neither, finite or not), we define

\[ \mathcal{B}(I) \equiv \sigma(\mathcal{O}(I)) \]

where \( \mathcal{O}(I) \) denotes the collection of all open sets contained in \( I \). The \( \sigma \)-field \( \sigma(\mathcal{O}(I)) \) is called the Borel \( \sigma \)-field on \( I \) and is denoted by \( \mathcal{B}(I) \).

This notion can be further extended: With \( A \) denoting a subset of \( \mathbb{R}^p \) for some positive integer \( p \), we write

\[ \mathcal{B}(A) \equiv \sigma(\mathcal{O}(A)) \]

where \( \mathcal{O}(A) \) denotes the collection of all open sets contained in \( A \). In particular, we have the following important definition.

**Definition 2.4.2** For each \( p = 1, 2, \ldots \), let \( \mathcal{O}(\mathbb{R}^p) \) denote the collection of all open sets contained in \( \mathbb{R}^p \). The \( \sigma \)-field \( \sigma(\mathcal{O}(\mathbb{R}^p)) \) is called the Borel \( \sigma \)-field on \( \mathbb{R}^p \) and is denoted by \( \mathcal{B}(\mathbb{R}^p) \). Thus,

\[ \mathcal{B}(\mathbb{R}^p) \equiv \sigma(\mathcal{O}(\mathbb{R}^p)). \]

The general definition of a Borel \( \sigma \)-field uses the collection of open sets as a generator for in higher-dimensions there are no intervals!

### 2.5 Mappings and measurability

Consider mappings \( g : S_a \rightarrow S_b \) and \( h : S_b \rightarrow S_c \) where \( S_a, S_b \) and \( S_c \) are arbitrary sets (possibly identical).

**Fact 2.5.1** Let \( \mathcal{B} \) be a collection of subsets of \( S_b \) (so \( \mathcal{B} \subseteq \mathcal{P}(S_b) \)). With

\[ g^{-1}(\mathcal{B}) \equiv \{ g^{-1}(F_b) : F_b \in \mathcal{B} \}, \]

it is always the case that

\[ g^{-1}(\sigma(\mathcal{B})) = \sigma(g^{-1}(\mathcal{B})) \]
Proof. The collection $g^{-1}(\sigma(B))$ is a $\sigma$-field on $S_a$, and it contains $g^{-1}(B)$, hence the inclusion $\sigma (g^{-1}(B)) \subseteq g^{-1} (\sigma(B))$. To establish the reverse inclusion, consider the collection $\mathcal{B}_g$ of subsets of $S_b$ defined by

$$\mathcal{B}_g \equiv \{ F_b \subseteq S_b : g^{-1}(F_b) \in \sigma (g^{-1}(B)) \}.$$ 

It is plain that $\mathcal{B}_g$ is a $\sigma$-field on $S_b$; as it obviously contains $\mathcal{B}$, it must also contain $\sigma (B)$ and the inclusion $g^{-1} (\sigma(B)) \subseteq \sigma (g^{-1}(B))$ follows.

Define the mapping $h \circ g : S_a \rightarrow S_c$ obtained by composing $g$ with $h$ through

$$(h \circ g)(\omega_a) = h(g(\omega_a)), \quad \omega \in S_a$$

Fact 2.5.2 If $\mathcal{C}$ be a collection of subsets of $S_c$ (so $\mathcal{C} \subseteq \mathcal{P}(S_c)$), then

$$(h \circ g)^{-1}(\mathcal{C}) = g^{-1} (h^{-1}(\mathcal{C})).$$

2.6 Borel mappings

Consider an arbitrary set $S$ equipped with a $\sigma$-field $\mathcal{S}$.

Definition 2.6.1 A mapping $g : S \rightarrow \mathbb{R}^p$ is said to be a Borel mapping if the conditions

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

are all satisfied where

$$g^{-1}(B) \equiv \{ s \in S : g(s) \in B \}.$$ 

This definition has some several simple but useful consequences.

Fact 2.6.1 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.

Proof. This is a simple consequence of the fact that

$$(h \circ g)^{-1}(B) = g^{-1} (h^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}^q).$$
Thus, if $B$ is an element of $\mathcal{B}(\mathbb{R}^p)$, then $h^{-1}(B)$ is an element of $\mathcal{B}(\mathbb{R}^p)$ by the Borel measurability of $h$, whence $(h \circ g)^{-1}(B)$ is an element of $\mathcal{S}$ by the Borel measurability of $g$.

The following fact is crucial for understanding the importance of probability distributions.

**Lemma 2.6.1** Let $\mathcal{G}$ denote a collection of subsets of $\mathbb{R}^p$ which generates the Borel $\sigma$-field $\mathcal{B}(\mathbb{R}^p)$, i.e.,

\begin{equation}
\mathcal{B}(\mathbb{R}^p) = \sigma(\mathcal{G}).
\end{equation}

The mapping $g : \mathcal{S} \to \mathbb{R}^p$ is a Borel mapping if and only if the weaker set of conditions

\begin{equation}
g^{-1}(E) \in \mathcal{S}, \quad E \in \mathcal{G}
\end{equation}

holds.

**Proof.** One implication is trivial since the conditions (2.19) constitute a subset of the conditions (2.17). To prove the reverse implication consider the collection $\mathcal{E}_g$ given by

\[ \mathcal{E}_g \equiv \{ E \subseteq \mathbb{R}^p : g^{-1}(E) \in \mathcal{S} \}. \]

The collection $\mathcal{E}_g$ is a $\sigma$-field on $\mathbb{R}^p$ because $\mathcal{S}$ is a $\sigma$-field on $S$. Under condition (2.19) we note the inclusion $\mathcal{G} \subseteq \mathcal{E}_g$, hence $\sigma(\mathcal{G}) \subseteq \mathcal{E}_g$ and the conditions (2.17) all hold since $\sigma(\mathcal{G}) = \mathcal{B}(\mathbb{R}^p)$ by assumption.

There are many generators known for the Borel $\sigma$-field $\mathcal{B}(\mathbb{R}^p)$. For instance, we have (2.18) with

- $\mathcal{G} = \mathcal{R}_{open}(\mathbb{R}^p)$ where $\mathcal{R}_{open}(\mathbb{R}^p)$ is the collection of all finite open rectangles, i.e.,

\[ \mathcal{R}_{open}(\mathbb{R}^p) \equiv \left\{ I_1 \times \ldots \times I_p, \quad I_k \in \mathcal{I}(\mathbb{R}), \quad k = 1, \ldots, p \right\} \]

where

\[ \mathcal{I}(\mathbb{R}) = \{ (a, b) : a, b \in \mathbb{R} \} \]
2.7. **CONSEQUENCES**

Use the following fact: For any open set $U$ in $\mathbb{R}^p$ there exists a countable family of open rectangles $\{R_i; i \in I\}$ in $\mathcal{R}_{\text{open}}(\mathbb{R}^p)$ with countable $I$ such that $U = \bigcup_{i \in I} R_i$. It is the analog of a similar fact encountered in one dimension.

- $G = \mathcal{R}_{\text{SW}}(\mathbb{R}^p)$ where $\mathcal{R}_{\text{SW}}(\mathbb{R}^p)$ is the collection of all closed Southwest rectangles, i.e.,

$$\mathcal{R}_{\text{SW}}(\mathbb{R}^p) \equiv \left\{ I_1 \times \ldots \times I_p, \quad \begin{array}{l} I_k = (-\infty, a_k] \\ a_k \in \mathbb{R} \\ k = 1, \ldots, p \end{array} \right\}.$$

### 2.7 Consequences

It follows from the discussion above that a mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if the seemingly weaker conditions

$$\left\{ s \in S : g(s) \in \prod_{i=1}^{p} (-\infty, a_k] \right\} \in \mathcal{S}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

all hold. Equivalently, a mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if

$$\left\{ s \in S : g_k(s) \leq a_k, \quad k = 1, \ldots, p \right\} \in \mathcal{S}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p$$

where it is understood that

$$g(s) = (g_1(s), \ldots, g_p(s)), \quad s \in S.$$

It is now plain that for each $k = 1, \ldots, p$, the component mapping $g_k : S \to \mathbb{R}$ is also a Borel mapping – Just take $a_\ell = \infty$ for all $\ell = 1, \ldots, k$ different from $k$. Conversely, since

$$\left\{ s \in S : g_k(s) \leq a_k, \quad k = 1, \ldots, p \right\} = \cap_{k=1}^{p} \left\{ s \in S : g_k(s) \leq a_k \right\}$$

for arbitrary $(a_1, \ldots, a_p)$ in $\mathbb{R}^p$, we see that the mapping $g : S \to \mathbb{R}^p$ is a Borel mapping if and only if each of the component mappings $g_1 : S \to \mathbb{R}, \ldots, g_p : S \to \mathbb{R}$ is a Borel mapping.

Most (if not all) mappings $\mathbb{R}^p \to \mathbb{R}^q$ encountered in applications are Borel mappings. Furthermore, any continuous mapping $\mathbb{R}^p \to \mathbb{R}^q$ can be shown to be a Borel mapping!
CHAPTER 2. MEASURABILITY
Chapter 3

Random variables and their distributions

So far we have been concerned with modeling the full random experiment $\mathcal{E}$, and this has led us to introduce the notion of a probability triple $(\Omega, \mathcal{F}, P)$. However, in many settings interest in not so much in the full model itself but rather in various numerical characteristics associated with the experiment. This is formalized through the notion of random variable (rv) which we now discuss.

### 3.1 Random variables

**Definition 3.1.1** Given a probability triple $(\Omega, \mathcal{F}, P)$, a mapping $X : \Omega \to \mathbb{R}^p$ is a random variable (rv) if

\[
X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}, \quad B \in \mathcal{B}(\mathbb{R}^p).
\]

In other words, the mapping $X : \Omega \to \mathbb{R}^p$ is a rv if it is a Borel mapping $X : \Omega \to \mathbb{R}^p$ – Here $S = \Omega$ and $S = \mathcal{F}$. We shall often write $[X \in B]$ in lieu of $X^{-1}(B)$ and $P[X \in B]$ for $P[[X \in B]]$.

In view of the earlier discussion the mapping $X : \Omega \to \mathbb{R}^p$ is a rv if and only if

\[
\{\omega \in \Omega : X_k(\omega) \leq a_k, \ k = 1, \ldots, p\} \in \mathcal{F}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p
\]

where it is understood that

\[
X(\omega) = (X_1(\omega), \ldots, X_p(\omega)), \quad \omega \in \Omega.
\]
This representation induces component mappings \( X_1, \ldots, X_p : \Omega \to \mathbb{R} \) defined in an obvious manner.

This last condition (3.2) can also be rewritten as

\[
\bigcap_{k=1}^{p} [X_k \leq a_k] \in \mathcal{F}, \quad (a_1, \ldots, a_p) \in \mathbb{R}^p.
\]

(3.3)

It is now plain that for each \( k = 1, \ldots, p \), the component mapping \( X_k : \Omega \to \mathbb{R} \) is also a rv – Just take \( a_\ell = \infty \) for all \( \ell = 1, \ldots, k \) different from \( k \). Here as well, we conclude to the following noteworthy fact.

**Fact 3.1.1** The mapping \( X : \Omega \to \mathbb{R}^p \) is a rv if and only if each of the component mappings \( X_1 : \Omega \to \mathbb{R}, \ldots, X_p : \Omega \to \mathbb{R} \) is a rv.

### 3.2 Probability distribution functions

Consider an \( \mathbb{R}^p \)-valued rv \( X : \Omega \to \mathbb{R}^p \).

**Definition 3.2.1** The probability distribution (function) of the rv \( X \) is the mapping \( F_X : \mathbb{R}^p \to [0, 1] \) defined by

\[
F_X(x) \equiv \mathbb{P} [X \in (-\infty, x_1] \times \ldots \times (-\infty, x_p)] = \mathbb{P} [X_1 \leq x_1, \ldots, X_p \leq x_p], \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p.
\]

(3.4)

with the notation \( X = (X_1, \ldots, X_p) \).

It turns out that there is as much probabilistic information in the probability distribution \( F_X : \mathbb{R}^p \to [0, 1] \) of the rv \( X \) as in

\[
\{ \mathbb{P} [X \in B] \mid B \in \mathcal{B}(\mathbb{R}^p) \}
\]

In fact, knowledge of \( F_X : \mathbb{R}^p \to \mathbb{R} \) allows a unique reconstruction of

\[
\{ \mathbb{P} [X \in B] \mid B \in \mathcal{B}(\mathbb{R}^p) \}.
\]

This is a consequence of Theorem

It is easy to see that the following properties when \( p = 1 \).

**Proposition 3.2.1** Given a rv \( X : \Omega \to \mathbb{R} \) with probability distribution function \( F_X : \mathbb{R} \to [0, 1] \), the following properties hold:
3.2. PROBABILITY DISTRIBUTION FUNCTIONS

(i) **Monotonicity:**

\[ F_X(x) \leq F_X(y), \quad x < y, \quad x, y \in \mathbb{R} \]

(ii) **Right-continuity:**

\[ \lim_{y \uparrow x} F_X(y) = F_X(x), \quad x \in \mathbb{R} \]

(iii) **Existence of a left limit:**

\[ \lim_{y \downarrow x} F_X(y) = F_X(x-) \quad \text{with} \quad \mathbb{P}[X = x] = F_X(y) - F_X(x-), \quad x \in \mathbb{R} \]

(iv) **Behavior at infinity:** Monotonically we have \( \lim_{x \to -\infty} F_X(x) = 0 \) and \( \lim_{x \to \infty} F_X(x) = 1 \).

**Proof.**

(i) The monotonicity of \( F_X \) is inherited from that of \( \mathbb{P} \) one we note that with \( x \) and \( y \) in \( \mathbb{R} \), we have \( [X \leq x] \subseteq [X \leq y] \) as soon as \( x < y \). Indeed, we have

\[
\mathbb{P}[X \leq y] = \mathbb{P}[X \leq x] + \mathbb{P}[x < X \leq y]
\]
or equivalently,

\[
F_X(y) - F_X(x) = \mathbb{P}[x < X \leq y].
\]

(ii) Pick \( x \) in \( \mathbb{R} \), and let \( \{y_n, \ n = 1, 2, \ldots\} \) denote a decreasing sequence in \( \mathbb{R} \) such that \( x \leq y_n \) for each \( n = 1, 2, \ldots \). By comments in (i) we have

\[
F_X(y_n) - F_X(x) = \mathbb{P}[x < X \leq y_n], \quad n = 1, 2, \ldots
\]
The sets \( [x < X \leq y_n], \ n = 1, 2, \ldots \) form a decreasing set sequence with

\[
\cap_{n=1}^\infty [x < X \leq y_n] = \emptyset
\]
and the conclusion \( \lim_{n \to \infty} \mathbb{P}[x < X \leq y_n] = 0 \) follows, whence \( \lim_{n \to \infty} F_X(y_n) = F_X(x) \).

(iii) Similarly, pick \( x \) in \( \mathbb{R} \), and let \( \{y_n, \ n = 1, 2, \ldots\} \) denote an increasing sequence in \( \mathbb{R} \) such that \( y_n \leq x \) for each \( n = 1, 2, \ldots \). By comments in (i) we have

\[
F_X(x) - F_X(y_n) = \mathbb{P}[y_n < X \leq x], \quad n = 1, 2, \ldots
\]
The sets \( [y_n < X \leq x], \ n = 1, 2, \ldots \) form an increasing set sequence with

\[
\cup_{n=1}^\infty [y < X \leq x_n] = [X = x]
\]
This time we get $\lim_{n \to \infty} \mathbb{P}[y_n < X \leq x] = \mathbb{P}[X = x]$, and the limit $\mathbb{P}[X = x]$ being independent of the sequence, the desired result follows.

(iv) Finally,

\[ \square \]

### 3.3 Probability distribution functions

Turning these properties into a definition we introduce the concept of a probability distribution (function).

**Definition 3.3.1** A probability distribution (function) on $\mathbb{R}$ is any mapping $F : \mathbb{R} \to [0, 1]$ such that

- **Monotonicity:**
  \[ F(x) \leq F(y), \quad x, y \in \mathbb{R} \]

- **Right-continuity:**
  \[ \lim_{y \downarrow x} F(y) = F(x), \quad x \in \mathbb{R} \]

- **Existence of left limits:**
  \[ \lim_{y \uparrow x} F(y) = F(x-) \quad x \in \mathbb{R} \]

- **Behavior at infinity:** Monotonically, we have
  \[ \lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1 \]

Obviously, if $X : \Omega \to \mathbb{R}$ is a rv, then its probability distribution function $F_X : \mathbb{R} \to [0, 1]$ is a probability distribution function. Conversely,

**Lemma 3.3.1** Any rv $X : \Omega \to \mathbb{R}$ generates a probability distribution function $F_X : \mathbb{R} \to [0, 1]$. Conversely, for any probability distribution function $F : \mathbb{R} \to [0, 1]$, there exists a probability triple $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ and a rv $X^* : \Omega^* \to \mathbb{R}$ defined on it such that

\[ \mathbb{P}^*[X^* \leq x] = F(x), \quad x \in \mathbb{R} \]
This is the basis of Monte-Carlo simulation. There exists a multi-dimensional analog to this fact to be discussed later on.

**Proof.** Take $\Omega^* = [0, 1]$, $\mathcal{F}^* = \mathcal{B}([0, 1])$ and $\mathbb{P}^* = \lambda$. Define the rv $X^* : \Omega^* \to \mathbb{R}$ by setting

$$X^*(\omega^*) = F^-(\omega^*), \quad \omega^* \in [0, 1]$$

where $F^- : [0, 1] \to [-\infty, \infty]$ is the generalized inverse of $F$ given by

$$F^-(u) = \inf (x \in \mathbb{R} : u \leq F(x)), \quad 0 \leq u \leq 1$$

with the understanding that $F^-(u) = \infty$ if the defining set is empty, i.e., $F(x) < u$ for all $x$ in $\mathbb{R}$.

### 3.4 Discrete distributions

A rv $X : \Omega \to \mathbb{R}^p$ is a discrete rv if there exists a countable subset $S \subseteq \mathbb{R}^p$ such that

$$\mathbb{P} [X \in S] = 1.$$

Note that

$$\mathbb{P} [X \in B] = \sum_{x \in S \cap B} \mathbb{P} [X = x], \quad B \in \mathcal{B}(\mathbb{R}^p).$$

It is often more convenient to characterize the distributional properties of the rv $X$ through its probability mass function (pmf) of the rv $X$ given by

$$p_X \equiv (p_X(x), \ x \in S)$$

with

$$p_X(x) = \mathbb{P} [X = x], \quad x \in S.$$

Well-known examples of discrete rvs (and of their distributions) include:

(i) Bernoulli $Ber(p)$ (with $0 \leq p \leq 1$)

(ii) Binomial $Bin(n; p)$ (with $n = 1, 2, \ldots$ and $0 \leq p \leq 1$)

(iii) Poisson $Poi(\lambda)$ (with $\lambda > 0$)

(iv) Geometric $Geo(p)$ (with $0 \leq p \leq 1$)
3.5 Absolutely continuous distributions

A rv \( X : \Omega \to \mathbb{R}^p \) is an (absolutely) continuous rv if there exists a Borel mapping \( f_X : \mathbb{R}^p \to \mathbb{R}_+^p \) such that

\[
P[X_i \leq x_i, \ i = 1, \ldots, p] = \int_{-\infty}^{x} f_X(\xi) d\xi, \quad x = (x_1, \ldots, x_p) \in \mathbb{R}^p.
\]

Well-known examples of continuous rvs (and of their distributions) include:

(i) Uniform \( \text{U}(a, b) \) (with \( a < b \) in \( \mathbb{R} \))

(ii) Exponential \( \text{Exp}(\lambda) \) (with \( \lambda > 0 \))

(iii) Gaussian \( \text{N}(m, \sigma^2) \) (with \( m, \sigma \) in \( \mathbb{R} \))

(iv) Cauchy \( \text{C}(m, a) \) (with \( m, a \) in \( \mathbb{R} \))

3.6 Properties of \( F_X \) when \( p \geq 1 \)

- Monotonicity needs to be modified and now reads

\[
P[x_k < X_k \leq y_k] \geq 0, \quad x_k, y_k \in \mathbb{R}, \quad k = 1, \ldots, p
\]

with the understanding that the quantity \( P[x_k < X_k \leq y_k] \) is expressed solely in terms of \( F_X : \mathbb{R}^p \to [0, 1] \).

- Right-continuity: With the understanding that \( y_k \downarrow x_k \) for each \( k = 1, \ldots, p \), we have

\[
\lim_{y_k \downarrow x_k} F_X(y) = F_X(x), \quad x \in \mathbb{R}^p
\]

- Existence of left limits: With the understanding that \( y_k \uparrow x_k \) for each \( k = 1, \ldots, p \), we have

\[
\lim_{y_k \uparrow x_k} F_X(y) = F_X(x) - \quad \text{with} \quad P[X = x] = F_X(y) - F_X(x), \quad x \in \mathbb{R}^p
\]
3.7. INDEPENDENCE OF RVS

- Behavior at infinity:

\[ \lim_{\min(x_k, k=1,...,p) \to -\infty} F_X(x) = 0 \]

and

\[ \lim_{\min(x_k, k=1,...,p) \to \infty} F_X(x) = 1 \]

3.7 Independence of rvs

Consider a collection of rvs \( \{X_i, i \in I\} \) which are all defined on some probability triple \((\Omega, \mathcal{F}, P)\). Assume that for each \( i \) in \( I \), the rv \( X_i \) is a \( \mathbb{R}^{p_i} \)-valued rv for some positive integer \( p_i \).

With \( I \) finite, we shall say that the rvs \( \{X_i, i \in I\} \) are mutually independent if for each selection of \( B_i \) in \( \mathcal{B}(\mathbb{R}^{p_i}) \) for each \( i \) in \( I \), the events

\[ \{[X_i \in B_i], i \in I\} \]

are mutually independent. It is easy to see that this is equivalent to requiring

\[ P[\cap_{i \in I}[X_i \in B_i]] = \prod_{i \in I} P[X_i \in B_i], \quad B_i \in \mathcal{B}(\mathbb{R}^{p_i}) \]

More generally, with \( I \) arbitrary (and possibly uncountable), the rvs \( \{X_i, i \in I\} \) are said to be mutually independent if for every finite subset \( J \subseteq I \), the rvs \( \{X_j, j \in J\} \) are mutually independent!

3.8 Product spaces

**Definition 3.8.1** Consider two arbitrary sets \( \Omega_a \) and \( \Omega_b \) (possibly identical). Let \( \mathcal{A} \) and \( \mathcal{B} \) denote non-empty collections of subsets of \( \Omega_a \) and \( \Omega_b \), respectively. While the collection \( \mathcal{A} \times \mathcal{B} \) is usually not a \( \sigma \)-field on \( \Omega_a \times \Omega_b \), even when \( \mathcal{A} \) and \( \mathcal{B} \) are themselves \( \sigma \)-fields, it can be shown that

\[ \sigma(\mathcal{A} \times \mathcal{B}) = \sigma(\sigma(\mathcal{A}) \times \sigma(\mathcal{B})). \]

Consider the probability triples \((\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \ldots, (\Omega_p, \mathcal{F}_p, \mathbb{P}_p)\). Their Cartesian product is the set \( \Omega \) defined by

\[ \Omega \equiv \Omega_1 \times \ldots \times \Omega_n. \]
We introduce the collection $\mathcal{F}_1 \times \ldots \times \mathcal{F}_p$ of subsets of $\Omega$ given by

$$\mathcal{F}_1 \times \ldots \times \mathcal{F}_p = \left\{ F_1 \times \ldots \times F_p, \quad F_k \in \mathcal{F}_k, \quad k = 1, \ldots, p \right\}.$$  

We write

$$\mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_p = \otimes_{k=1}^p \mathcal{F}_k = \sigma (\mathcal{F}_1 \times \ldots \times \mathcal{F}_p).$$

Note that

$$\sigma (\mathcal{F}_1 \times \ldots \times \mathcal{F}_p) = \sigma (\sigma (\mathcal{F}_1) \times \ldots \times \sigma (\mathcal{F}_p)).$$

The product probability measure $\mathbb{P}$ is defined on $\otimes_{k=1}^p \mathcal{F}_k$ as follows: For any rectangle

$$R = F_1 \times \ldots \times F_p, \quad F_k \in \mathcal{F}_k, \quad k = 1, \ldots, p$$

set

$$(3.5) \quad \mathbb{P} [R] = \prod_{k=1}^p \mathbb{P}_k [F_k].$$

So far, $\mathbb{P}$ is defined only on $\mathcal{F}_1 \times \ldots \times \mathcal{F}_p$. However, Measure Theory guarantees that there exists a unique probability measure on the $\sigma$-field

$$\sigma (\sigma (\mathcal{F}_1) \times \ldots \times \sigma (\mathcal{F}_p))$$

such that (3.5) holds.

An important modeling fact Under $\mathbb{P}$, the events

$$\begin{align*}
E_1 &= A_1 \times \Omega_2 \times \ldots \times \Omega_p \\
E_2 &= \Omega_1 \times A_2 \times \ldots \times \Omega_p \\
&\vdots \\
E_p &= \Omega_1 \times \Omega_2 \times \ldots \times A_p
\end{align*}$$

are mutually independent with

$$\mathbb{P} [E_k] = \mathbb{P}_k [A_k], \quad k = 1, \ldots, p.$$
3.9 Taking limits

Consider the sequence of $\mathbb{R}$-valued rvs $\{X_n, n = 1, 2, \ldots\}$ which are all defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. The following mappings $\Omega \to [-\infty, \infty]$ are rvs in the extended sense:

- The supremum mapping $\Omega \to [-\infty, \infty]$ defined by
  \[ \omega \to \sup_{n \geq 1} X_n(\omega), \quad \omega \in \Omega \]

- The infimum mapping $\Omega \to [-\infty, \infty]$ defined by
  \[ \omega \to \inf_{n \geq 1} X_n(\omega), \quad \omega \in \Omega \]

- The limsup mapping $\Omega \to [-\infty, \infty]$ defined by
  \[ \omega \to \lim_{n \to \infty} \sup X_n(\omega), \quad \omega \in \Omega \]

- The liminf mapping $\Omega \to [-\infty, \infty]$ defined by
  \[ \omega \to \lim_{n \to \infty} \inf X_n(\omega), \quad \omega \in \Omega \]

It follows that

\[ \Omega^* \equiv \left[ \lim_{n \to \infty} \inf X_n = \lim_{n \to \infty} \sup X_n \right] \in \mathcal{F} \]

and on $\Omega^*$, it holds that $\lim_{n \to \infty} X_n$ exists (possibly as an element in $[-\infty, \infty]$), and is the common value assumed by $\liminf_{n \to \infty} X_n$ and $\limsup_{n \to \infty} X_n$.

When $\mathbb{P}[\Omega^*] = 1$ it is customary to say that the sequence $\{X_n, n = 1, 2, \ldots\}$ converges almost surely (a.s.) (under $\mathbb{P}$), and we write

\[ \lim_{n \to \infty} X_n \quad \mathbb{P}\text{-a.s.} \]

In that case, for any rv $X: \Omega \to \mathbb{R}$ such that

\[ X(\omega) = \lim_{n \to \infty} X_n(\omega), \quad \omega \in \Omega^* \]

we shall write

\[ \lim_{n \to \infty} X_n = X \quad \mathbb{P}\text{-a.s.} \]
CHAPTER 3. RANDOM VARIABLES AND THEIR DISTRIBUTIONS

Such a rv $X$ always exists when $\mathbb{P}[\Omega^*] = 1$ but is not unique. Existence is immediate since we can always take

$$X(\omega) \equiv \begin{cases} 
\lim \inf_{n \to \infty} X_n(\omega) = \lim \sup_{n \to \infty} X_n(\omega) & \text{if } \omega \in \Omega^* \\
Z(\omega) & \text{if } \omega \notin \Omega^* 
\end{cases}$$

where $Z : \Omega \to \mathbb{R}$ is some arbitrary rv, and non-uniqueness is obvious.

3.10 Simple rvs

A rv $X : \Omega \to \mathbb{R}$ is a simple variable if

$$X = \sum_{k \in I} a_k 1[A_k]$$

where (i) $I$ is a finite index set, (ii) $\{a_k, k \in I\}$ are scalars (not necessarily distinct) and (iii) the subsets $\{A_k, k \in I\}$ form an $\mathcal{F}$-partition of $\Omega$, i.e., the subsets $\{A_k, k \in I\}$ are all in $\mathcal{F}$ with

$$\bigcup_{k \in I} A_k = \Omega \quad \text{and} \quad A_k \cap A_\ell = \emptyset \quad k \neq \ell \quad k, \ell \in I.$$ 

This representation is not necessarily unique. In many arguments it is customary to assume that the values $\{a_k, k \in I\}$ are distinct scalars and that the events $\{A_k, k \in I\}$ forming the $\mathcal{F}$-partition are all non-empty, in which case $\{X(\omega), \omega \in \Omega\} = \{a_k, k \in I\}$ and

$$A_k = [X = a_k], \quad k \in I.$$ 

We refer to this representation as the generic representation of the simple rv. There is no loss of generality in using the generic representation as will shortly become apparent.

**Lemma 3.10.1 Fact:** For any rv $X : \Omega \to \mathbb{R}_+$, there exists a monotonically increasing sequence of simple rvs $\{X_n, n = 1, 2, \ldots\}$ such that

$$X_n \leq X_{n+1} \leq X, \quad n = 1, 2, \ldots$$

and

$$\lim_{n \to \infty} X_n = X.$$
For instance, for each $n = 1, 2, \ldots$, define the simple rv $X_n : \Omega \to \mathbb{R}_+$ by

\[
X_n = \sum_{m=0}^{n-1} \sum_{k=0}^{2^n-1} (m + k2^{-n}) \mathbf{1} [m + k2^{-n} < X \leq m + (k + 1)2^{-n}]
\]
Part II

APENDICES
Chapter 4

Limits in $\mathbb{R}$

We begin with a few standard definitions. We refer to a mapping $a : \mathbb{N}_0 \to \mathbb{R}$ as a ($\mathbb{R}$-valued) sequence; sometimes we also use the notation $\{a_n, \ n = 1, 2, \ldots\}$.

A sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges to $a^*$ in $\mathbb{R}$ if for every $\varepsilon > 0$, there exists an integer $n^*(\varepsilon)$ such that

$$|a_n - a^*| \leq \varepsilon, \quad n \geq n^*(\varepsilon).$$

(4.1)

We shall write $\lim_{n \to \infty} a_n = a^*$, and refer to the scalar $a^*$ as the limit of the sequence.

Sometimes it is desirable to make sense of situations where values of the sequence become either unboundedly large or unboundedly negative, in which case we shall write $\lim_{n \to \infty} a_n = \infty$ and $\lim_{n \to \infty} a_n = -\infty$, respectively. A precise definition of such occurrences is as follows: We write $\lim_{n \to \infty} a_n = \infty$ to signify that for every $M > 0$, there exists a finite integer $n^*(M)$ in $\mathbb{N}_0$ such that

$$a_n > M, \quad n \geq n^*(M).$$

(4.2)

It is natural to define $\lim_{n \to \infty} a_n = -\infty$ if $\lim_{n \to \infty} (-a_n) = \infty$.

If there exists $a^*$ in $\mathbb{R} \cup \{\pm \infty\}$ such that $\lim_{n \to \infty} a_n = a^*$, we shall simply say that the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges or is convergent (without any reference to its limit). Sometimes we shall also say that the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ converges in $\mathbb{R}$ to indicate that the limit $a^*$ is an element of $\mathbb{R}$ (thus finite).

Applying the definition (4.1) requires that the limit be known. Often this information is not available, and yet the need remains to check whether the sequence
converges. The notion of *Cauchy sequence*, which is instrumental in that respect, is built around the following observation: If the sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) converges to \( a^* \) in \( \mathbb{R} \), then for every \( \varepsilon > 0 \), there exists a finite integer \( n^*(\varepsilon) \) such that (4.1) holds, hence for \( n, m \geq n^*(\varepsilon) \) we have

\[
|a_n - a_m| \leq |a_n - a^*| + |a^* - a_m| \leq \varepsilon + \varepsilon = 2\varepsilon.
\]

This observation is turned into the following definition.

A sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) is said to be a *Cauchy sequence* if for every \( \varepsilon > 0 \), there exists an integer \( n^*(\varepsilon) \) such that

\[
|a_n - a_m| \leq \varepsilon, \quad m, n \geq n^*(\varepsilon).
\] (4.3)

As observed earlier, a convergent sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) in \( \mathbb{R} \) is always a Cauchy sequence. It is a deep fact concerning the topological properties of \( \mathbb{R} \) that being a Cauchy sequence is sufficient for convergence of the sequence in \( \mathbb{R} \).

**Theorem 4.0.1** (*Cauchy criterion*) A sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) is convergent in \( \mathbb{R} \) if and only if it is a Cauchy sequence.

This provides a criterion for convergence which does not require knowledge of the limit.

### 4.1 Two important facts

In addition to the Cauchy convergence criterion, here are two facts that are often found useful in studying convergence.

A sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) is said to be *non-decreasing* (resp. *non-increasing*) if

\[
a_n \leq a_{n+1} \quad \text{(resp. } a_{n+1} \leq a_n)\] , \quad n = 1, 2, \ldots
\]

A *monotone* sequence is a sequence that is either non-decreasing or non-increasing.

**Theorem 4.1.1** A monotone sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) always converges and we have

\[
\lim_{n \to \infty} a_n = \sup (a_n, \ n = 1, 2, \ldots) \quad \text{(resp. } \lim_{n \to \infty} a_n = \inf (a_n, \ n = 1, 2, \ldots))
\]

if the sequence is non-decreasing (resp. non-increasing).
Consider a sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \). A subsequence of the sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \) is any sequence of the form \( N_0 \rightarrow \mathbb{R} : k \rightarrow a_{n_k} \) where
\[
n_k < n_{k+1}, \quad k = 1, 2, \ldots
\]
This forces \( \lim_{k \to \infty} n_k = \infty \).

The sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \) is said to be bounded if there exists some \( B > 0 \) such that
\[
\sup (|a_n|, n = 1, 2, \ldots) \leq B.
\]

**Theorem 4.1.2 (Bolzano-Weierstrass)** For any bounded sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \), there exists a convergent subsequence \( N_0 \rightarrow \mathbb{R} : k \rightarrow a_{n_k} \) with \( \lim_{k \to \infty} a_{n_k} = a^* \) for some \( a^* \) in \( \mathbb{R} \).

### 4.2 Accumulation points

Since not all sequences converge, it is important to understand how non-convergence occurs.

An accumulation point for the sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \) is defined as any \( a^* \) in \( \mathbb{R} \cup \{\pm \infty\} \) such that
\[
\lim_{k \to \infty} a_{n_k} = a^*
\]
for some subsequence \( N_0 \rightarrow \mathbb{R} : k \rightarrow a_{n_k} \).

A convergent sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \) has exactly one accumulation point, namely its limit. In fact, were the sequence not convergent, it must necessarily have distinct accumulation points (in \( \mathbb{R} \cup \{\pm \infty\} \)), in which case there is a smallest and a largest. The next definition formalizes this observation. Given a sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \), the quantities
\[
\bar{A} = \limsup_{n \to \infty} A_n = \inf_{n \geq 1} \left( \sup_{m \geq n} a_m \right)
\]
and
\[
\underline{A} = \liminf_{n \to \infty} A_n = \sup_{n \geq 1} \left( \inf_{m \geq n} a_m \right)
\]
are known as the limsup and liminf of the sequence \( a : \mathbb{N}_0 \rightarrow \mathbb{R} \).
The following notation is found to be convenient when using liminf and limsup quantities: For each \( n = 1, 2, \ldots \), we define the quantities

\[
\bar{A}_n = \sup_{m \geq n} a_m \quad \text{and} \quad \underline{A}_n = \inf_{m \geq n} a_m
\]

(4.4)

Note that \( \underline{A}_n \leq \bar{A}_n \), and the sequences \( n \to \bar{A}_n \) and \( n \to \underline{A}_n \) are non-increasing and non-decreasing, respectively. Therefore, \( \bar{A} = \lim_{n \to \infty} \bar{A}_n \) and \( \underline{A} = \lim_{n \to \infty} \underline{A}_n \) both exist, but are possibly infinite. Moreover, we always have \( \underline{A} \leq \bar{A} \).

**Theorem 4.2.1** Consider a sequence \( a : \mathbb{N}_0 \to \mathbb{R} \). If it converges to \( a^\ast \), then \( \bar{A} = \underline{A} = a^\ast \). Conversely, if \( \bar{A} = \underline{A} = a^\ast \) for some \( a^\ast \) in \( \mathbb{R} \cup \{\pm \infty\} \), then the sequence converges to \( a^\ast \).

Note that if \( a, b : \mathbb{N}_0 \to \mathbb{R} \) are two sequences such that

\[ a_n \leq b_n, \quad n = 1, 2, \ldots \]

then \( \bar{A} \leq \bar{B} \) and \( \underline{A} \leq \underline{B} \). The following arguments will often be made on the basis of this observation: Consider a sequence \( \{p_n, n = 1, 2, \ldots\} \) where for each \( n = 1, 2, \ldots, p_n \) is the probability of some event so that

\[
0 \leq p_n \leq 1, \quad n = 1, 2, \ldots
\]

(4.5)

If we show that

\[
1 \leq \lim_{n \to \infty} \inf p_n,
\]

(4.6)

then we necessarily have convergence of the sequence with \( \lim_{n \to \infty} p_n = 1 \): Indeed, we always have \( \limsup_{n \to \infty} p_n \leq 1 \) as a result of (4.5), whence

\[
\lim_{n \to \infty} \inf p_n = \limsup_{n \to \infty} p_n = 1
\]

upon using (4.6). In a similar vein, if we show \( \limsup_{n \to \infty} p_n = 0 \), then we necessarily have convergence of the sequence with \( \lim_{n \to \infty} p_n = 0 \).

### 4.3 Cesaro convergence

With any sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) we associate the **Cesaro sequence** \( \bar{a} : \mathbb{N}_0 \to \mathbb{R} \) given by

\[
\bar{a}_n = \frac{1}{n} (a_1 + \ldots + a_n), \quad n = 1, 2, \ldots
\]
4.3. CESARO CONVERGENCE

Theorem 4.3.1 (Cesaro convergence) If the sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) converges to \( a^* \), then the Cesaro sequence \( \bar{a} : \mathbb{N}_0 \to \mathbb{R} \) also converges with limit \( a^* \).

Proof. First we assume the convergent sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) to have a finite limit \( a^* \) in \( \mathbb{R} \). Note that
\[
\bar{a}_n - a^* = \frac{1}{n} \sum_{k=1}^{n} (a_k - a^*), \quad n = 1, 2, \ldots
\]
Now, for every \( \varepsilon > 0 \), there exists an integer \( n^*(\varepsilon) \) such that
\[
|a_n - a^*| \leq \frac{\varepsilon}{2}, \quad n \geq n^*(\varepsilon).
\]
On that range, with \( B(\varepsilon) = \sum_{k=1}^{n^*(\varepsilon)} |a_k - a^*| \), we have
\[
|\bar{a}_n - a^*| \leq \frac{1}{n} \sum_{k=1}^{n} |a_k - a^*| \leq \frac{1}{n} \sum_{k=1}^{n^*(\varepsilon)} |a_k - a^*| + \frac{1}{n} \sum_{k=n^*(\varepsilon)+1}^{n} |a_k - a^*| \leq \frac{B(\varepsilon)}{n} + \frac{n - n^*(\varepsilon)}{n} \cdot \varepsilon
\]
(4.7)

Since \( \lim_{n \to \infty} \frac{1}{n} = 0 \), for every \( \varepsilon > 0 \), there exists a finite integer \( n^{**}(\varepsilon) \) such that
\[
\frac{1}{n} < \frac{\varepsilon}{B(\varepsilon)}, \quad n \geq n^{**}(\varepsilon).
\]
Just take \( n^{**}(\varepsilon) = \lceil \frac{B(\varepsilon)}{\varepsilon} \rceil \). As a result,
\[
|\bar{a}_n - a^*| \leq \varepsilon + \varepsilon = 2\varepsilon, \quad n \geq \max(n^*(\varepsilon), n^{**}(\varepsilon))
\]
and the proof is now complete since \( \varepsilon \) is arbitrary. We leave it as an exercise to show the result when \( a^* = \pm \infty \).
CHAPTER 4. LIMITS IN $\mathbb{R}$

However, the converse is not true: Take the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ given by

$$a_n = (-1)^n, \quad n = 1, 2, \ldots$$

This sequence does not converge and yet $\lim_{n \to \infty} \bar{a}_n = 0$. This example nicely illustrate the smoothing effect of averaging. It might be tempting to conjecture that such averaging always produces a convergent sequence. However, this is not so as the following example shows: Consider the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ given by

$$a_n = (-1)^k, \quad 2^k \leq n < 2^{k+1}, \quad k = 0, 1, \ldots$$

It is plain that $\lim \inf_{n \to \infty} a_n = -1$ while $\lim \sup_{n \to \infty} a_n = 1$, and so the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ does not converge. However, it is also not Cesaro convergent.

4.4 Series

Starting with a sequence $a : \mathbb{N}_0 \to \mathbb{R}$, we define the partial sums

$$s_n = a_1 + \ldots + a_n, \quad n = 1, 2, \ldots$$

where $s_n$ is known as the $n^{th}$ partial sum. We refer to the sequence $s : \mathbb{N}_0 \to \mathbb{R} : n \to s_n$ as the sequence of partial sums associated with the sequence $a : \mathbb{N}_0 \to \mathbb{R}$.

It is customary to say that the series $\sum_{n=1}^{\infty} a_n$ converges if the sequence $s : \mathbb{N}_0 \to \mathbb{R} : n \to s_n$ converges to some $s^*$ in $\mathbb{R}$, in which case we often write $\sum_{n=1}^{\infty} a_n$ as its limit.

This amounts to the following: For every $\varepsilon > 0$ there exists a finite integer $n^*(\varepsilon)$ such that

$$|s_n - s^*| < \varepsilon, \quad n \geq n^*(\varepsilon).$$

The series $s : \mathbb{N}_0 \to \mathbb{R}$ is said to be **absolutely convergent** if the series associated with the sequence of absolute values $\mathbb{N}_0 \to \mathbb{R}^+ : n \to |a_n|$ does itself converge in $\mathbb{R}$.

A series which is absolutely convergent is also convergent in the usual sense since

$$\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k|, \quad m = n + 1, \ldots \quad n = 1, 2, \ldots$$
However, the converse is not true as is easily seen through the example

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

A series which is convergent in the usual sense but not absolutely convergent is said to be **conditionally** convergent.

When the sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) assumes only non-negative values, i.e., \( a_n \geq 0 \) for all \( n = 1, 2, \ldots \), then the sequence \( s : \mathbb{N}_0 \to \mathbb{R}_+ \) of partial sums is non-decreasing, so that \( \lim_{n \to \infty} s_n \) always exists, possibly infinite. When this limit is finite, it is easy to establish the following fact.

**Lemma 4.4.1** For any sequence \( a : \mathbb{N}_0 \to \mathbb{R} \) whose sequence of partial sums converges in \( \mathbb{R} \), we have \( \lim_{n \to \infty} a_n = 0 \)

**Proof.** Since the sequence of partial sums \( s : \mathbb{N}_0 \to \mathbb{R} \) converges in \( \mathbb{R} \), it is a Cauchy sequence: For every \( \varepsilon > 0 \), there exists a finite integer \( n^*(\varepsilon) \) such that

\[ |s_n - s_m| \leq \varepsilon, \quad n, m \geq n^*(\varepsilon). \]

Selecting \( m = n + 1 \) with \( n \geq n^*(\varepsilon) \), we get \( |a_{n+1}| = |s_n - s_{n+1}| \leq \varepsilon \) whenever \( n \geq n^*(\varepsilon) \), and the conclusion \( \lim_{n \to \infty} a_n = 0 \) follows. \[ \blacksquare \]

Many tests exist to check the convergence of series. The most basic one is the **Comparison Test** given next.

**Theorem 4.4.1 (Comparison Test) Consider two sequences \( a, b : \mathbb{N}_0 \to \mathbb{R}_+ \) such that**

\[ 0 \leq a_n \leq b_n, \quad n = 1, 2, \ldots \]

If \( \sum_{n=1}^\infty b_n \) converges in \( \mathbb{R} \), then \( \sum_{n=1}^\infty a_n \) also converges in \( \mathbb{R} \) with

\[ 0 \leq \sum_{n=1}^\infty a_n \leq \sum_{n=1}^\infty b_n. \]

On the other hand, if \( \sum_{n=1}^\infty a_n = \infty \), then we necessarily have \( \sum_{n=1}^\infty b_n = \infty \).
Geometric series play a pivotal role in determining the convergence of series through the Comparison Test. The geometric series with reason $\rho$ is the series associated with the sequence $a : \mathbb{N}_0 \to \mathbb{R}$ given by

$$a_n = \rho^n, \quad n = 1, 2, \ldots$$

It well known that

$$s_n = a_1 + \ldots + a_n = \begin{cases} \frac{\rho}{1-\rho} (1 - \rho^n) & \text{if } \rho \neq 1 \\ n & \text{if } \rho = 1 \end{cases}$$

Therefore,

$$\lim_{n \to \infty} s_n = \frac{\rho}{1-\rho} \quad \text{if } |\rho| < 1.$$  

This observation constitutes the basis for two criteria for convergence of series, namely the criteria of Cauchy and d’Alembert, also known as the Root Test and Ratio Test, respectively.

**Theorem 4.4.2 (Ratio Test)** Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. Assume that the limit

$$\lim_{n \to \infty} \frac{|a_{n+1}|}{|a_n|} = R$$

exists (possibly infinite). Then, $\sum_{n=1}^{\infty} |a_n| < \infty$ if $R < 1$ and $\sum_{n=1}^{\infty} |a_n| = \infty$ if $1 < R$.

**Theorem 4.4.3 (Root Test)** Consider a sequence $a : \mathbb{N}_0 \to \mathbb{R}$. Assume that the limit

$$\lim_{n \to \infty} \sqrt[n]{|a_n|} = R$$

exists. Then, $\sum_{n=1}^{\infty} |a_n| < \infty$ if $R < 1$ and $\sum_{n=1}^{\infty} |a_n| = \infty$ if $1 < R$.

### 4.5 Power series

In a number of places we shall need to understand the behavior of series that belong to the class of *power series*. With any sequence $a : \mathbb{N}_0 \to \mathbb{R}$ we associate the *formal* power series

$$\sum_{n=1}^{\infty} a_n z^n, \quad z \in \mathbb{C}.$$
A natural question arises as to when such formal series are in fact convergent. In particular, we define the domain of convergence of the power series as the set $C$ given by

$$C = \{ z \in \mathbb{C} : \sum_{n=0}^{\infty} |a_n|z^n < \infty \}.$$ 

This region is determined by the asymptotic behavior of the sequence $a : \mathbb{N}_0 \to \mathbb{R}$. This is the content of the following well-known result which is a consequence of the Root Test (applied to the sequence \{a_n z^n, n = 0, 1, \ldots\}).

**Theorem 4.5.1** With

$$R = \limsup_{n \to \infty} \sqrt[n]{|a_n|},$$

we have $\sum_{n=1}^{\infty} |a_n||z|^n < \infty$ if $|z| < R^{-1}$ and $\sum_{n=1}^{\infty} |a_n||z|^n = \infty$ if $R^{-1} < |z|$. The open disk $\{ z \in \mathbb{C} : |z| < R^{-1} \}$ is therefore contained in $C$, and $R^{-1}$ is known as the radius of convergence of the power series.