# CHAPTER 3
# Preliminaries from Probability Theory

La théorie des probabilités n'est autre que le sens commun fait calcul
Pierre Simon Laplace, 1819

This chapter recalls a few basic notions from probability theory and introduces some notation that is used throughout this thesis.

The following definition of sum allows one to work with sums over arbitrary sets. For countable sums, the notion coincides with the usual one.

## 3.1 Probability Distributions

**Definition 3.1.1** Let $\mathcal{I}$ be an index set and let $x_i \in [0, \infty]$ for all $i \in \mathcal{I}$. Define $\sum_{i \in \mathcal{I}} x_i$ by

1. $\sum_{i \in \emptyset} x_i = 0$,
2. $\sum_{i \in \mathcal{I}} x_i = x_{i_1} + x_{i_2} + x_{i_3} + \cdots + x_{i_n}$, if $\mathcal{I} = \{i_1, i_2, i_3, \ldots, i_n\}$ is a finite set with $n > 0$ elements,
3. $\sum_{i \in \mathcal{I}} x_i = \mathrm{supr}\{\sum_{i \in \mathcal{J}} x_i \mid \mathcal{J} \subseteq \mathcal{I} \text{ is finite}\}$, if $\mathcal{I}$ is infinite.

Here supr $X$ denotes the supremum of a set $X$. Notice that $\sum_{i \in \mathbb{N}} x_i = \sum_{i=0}^{\infty} x_i$ because, due to the fact that $x_i \geq 0$, the summation order is irrelevant. It is also easy to see that, if $\sum_{i=0}^{\infty} x_i$ is finite, then there are countably[1] many nonzero elements $x_i$.

**Definition 3.1.2** A *discrete probability distribution*,[2] or simply a *probability distribution*, over a set $X$ is a function $\mu : X \to [0, 1]$ such that $\sum_{x \in X} \mu(x) = 1$. We write $\mathrm{supp}(\mu) = \{x \in X \mid \mu(x) > 0\}$. It follows immediately from the definitions that this is a countable set. We denote the set of all probability distributions over $X$ by $\mathrm{Distr}(X)$.

We denote a probability distribution $\mu$ on a countable domain by enumerating it as a set of pairs. So, if $\mathrm{Dom}(\mu) = \{x_1, x_2 \ldots\}$ then we denote $\mu$ by $\{x_1 \mapsto \mu(x_1), x_2 \mapsto \mu(x_2) \ldots\}$. If the domain of $\mu$ is known from the context, then we often leave out elements of probability

[1] In our terminology, the countable sets subsume the finite sets.

[2] Here, we deviate from the terminology used in classical probability theory. In that setting, the outcomes of probabilistic experiments are described by random variables. The probability distribution of a random variable $Z$ is the function $z \mapsto \mathbf{P}[Z \leq z]$. For discrete random variables, the function $z \mapsto \mathbf{P}[Z = z]$ is a probability distribution in our sense and is called the *probability density function* or *probability mass function* of $Z$.

zero. For instance, the probability distribution assigning probability one to an element $x \in X$ is denoted by $\{x \mapsto 1\}$, irrespective of $X$. This distribution is called the *Dirac distribution* over $x$. The *uniform distribution* over a finite set with $n > 0$ elements, say $\{x_1, \dots, x_n\}$, is given by $\{x_1 \mapsto \frac{1}{n}, \dots, x_n \mapsto \frac{1}{n}\}$.

**Definition 3.1.3** Let $X$ and $Y$ be sets, $\mu \in \mathrm{Distr}(X)$ and $\nu \in \mathrm{Distr}(Y)$. The *product* of $\mu$ and $\nu$, notation $\mu \times \nu$, is the probability distribution $\kappa : X \times Y \to [0,1]$ such that $\kappa(x, y) = \mu(x) \cdot \nu(y)$.