# Contents

# 1 Probability

## 1.1 Probability Spaces

In an elementary level, we have been viewing probability as the quotient between the number of desired outcomes and the number of all possible outcomes. This definition, though intuitive, is not very solid when it comes to an infinite sample space. In this introductory chapter, we would establish the theories of probability using a more modern and rigorous structure.

**Definition 1.1.1 ▸ Set Algebra**

Let $X$ be a set. A **set algebra** over $X$ is a family $\mathcal{F} \subseteq \mathcal{P}(X)$ such that

- $X \backslash F \in \mathcal{F}$ for all $F \in \mathcal{F}$ (closed under complementation);
- $X \in \mathcal{F}$;
- $X_1 \cup X_2 \in \mathcal{F}$ for any $X_1, X_2 \in \mathcal{F}$ (closed under binary union).

There are several immediate implications from the above definition.

First, by closure under complementation, we know that an algebra over any set $X$ must contain the empty set.

Second, by De Morgan's Law, one can easily check that if the first 2 axioms hold, the closure under binary union is equivalent to

- $X_1 \cap X_2 \in \mathcal{F}$ for any $X_1, X_2 \in \mathcal{F}$;
- $\bigcup_{i=1}^{n} X_i \in \mathcal{F}$ for any $X_1, X_2, \cdots, X_n \in \mathcal{F}$ for all $n \in \mathbb{N}$;
- $\bigcap_{i=1}^{n} X_i \in \mathcal{F}$ for any $X_1, X_2, \cdots, X_n \in \mathcal{F}$ for all $n \in \mathbb{N}$.

$(X, \mathcal{F})$ is known as a *field of sets*, where the elements of $X$ are called *points* and those of $\mathcal{F}$, *complexes* or *admissible sets* of $X$.

In probability theory, what we are interested in is a special type of set algebras known as *$\sigma$-algebras*.

### Definition 1.1.2 ▸ $\sigma$-Algebra

A **$\sigma$-Algebra** over a set $A$ is a non-empty set algebra over $A$ that is closed under countable union.

Of course, by the same argument as above, we known that any $\sigma$-algebra is closed under countable intersection as well.

Now, as we all know, we can take some set $\Omega$ as a *sample space* and denote an *event* by some subset of $\Omega$. Roughly speaking, we could now define the probability of an event $E \subseteq \Omega$ as the ratio between the sets' volumes. The remaining question now is: how do we define the volume of a set properly?

### Definition 1.1.3 ▸ Measure

Let $X$ be a set and $\Sigma$ be a $\sigma$-algebra over $X$. A **measure** over $\Sigma$ is a function

$$\mu : \Sigma \to \mathbb{R} \cup \{-\infty, +\infty\}$$

such that

- $\mu(E) \geq 0$ for all $E \in \Sigma$ (non-negativity);
- $\mu(\varnothing) = 0$;
- $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$ for any countable collection of pairwise disjoint elements of $\Sigma$ (countable additivity or $\sigma$-additivity).

The triple $(X, \Sigma, \mu)$ is known as a **measure space** and the pair $(X, \Sigma)$, a **measurable space**.

One thing to note here is that if at least one $E \in \Sigma$ has a finite measure, then $\mu(\varnothing) = 0$ is automatically guaranteed for obvious reasons.

### Definition 1.1.4 ▸ Probability Space

Let $\Omega$ be a sample space and $\mathcal{F}$ be a $\sigma$-algebra over $\Omega$. A **probability space** is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P} : \mathcal{F} \to [0, 1]$, known as a **probability measure**, is such that $\mathbb{P}(\Omega) = 1$.

Obviously, the above definition immediately guarantees that

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
2. $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $\mathbb{P}(A) \subseteq \mathbb{P}(A)$;
3. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

The third result follows from a direct application of the principle of inclusion and exclusion.

By induction, one can easily check that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i)$$

for any finitely many events. The following proposition extends this result to countable collections of events:

### Proposition 1.1.5 ▶ Union Bound of Countable Collections of Events

*Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $E_1, E_2, \cdots, E_n, \cdots \in \mathcal{F}$ is any countable sequence of events, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

*Proof.* Define $F_1 := E_1$ and $F_k := E_k \setminus \bigcup_{i=1}^{k-1} E_i$ for $k \geq 2$. Clearly, the $F_i$'s are pairwise disjoint. By Definition 1.1.2, the $F_i$'s are elements of $\mathcal{F}$. Note that $\mathbb{P}(F_i) \leq \mathbb{E}_{\mathbb{i}}$ for all $i \in \mathbb{N}^+$, so

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i\right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(F_i) \\ &\leq \sum_{i=1}^{\infty} \mathbb{P}(E_i).\end{aligned}$$

□

Next, we will introduce the notion of *random variables* formally. For this purpose, we first establish the notion of a *Borel algebra*.

### Definition 1.1.6 ▶ Borel Algebra

Let $X$ be a topological space. A **Borel set** on $X$ is a set which can be formed via countable union, countable intersection and relative complementation of open sets in $X$. The smallest $\sigma$-algebra over $X$ containing all Borel sets on $X$ is known as the **Borel algebra** over $X$.

Clearly, the Borel algebra over $X$ contains all open sets in $X$ according to the above axioms from Definition 1.1.2. This helps us define the following:

**Definition 1.1.7 ▸ Random Variable**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{X}, \mathcal{B})$ be a measurable space where $\mathcal{B}$ is the Borel algebra over $\mathcal{X}$. A **random variable** is a function $X: \Omega \to \mathcal{X}$ such that

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

for all $B \in \mathcal{B}$.

*Remark.* Rigorously, such a random variable $X$ is a *measurable function* or *measurable mapping* from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{B})$.

The probability measure $\mathbb{P}$ thus induces a probability measure $P_X$ over $(\mathcal{X}, \mathcal{B})$.

**Definition 1.1.8 ▸ Distribution**

Let $X: \Omega \to \mathcal{X}$ be a random variable over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{B}$ be the Borel algebra over $\mathcal{X}$, the **distribution** of $X$ is the probability measure $P_X$ on $(\mathcal{X}, \mathcal{B})$ given by

$$P_X(B) = \{\omega \in \Omega : \mathbb{X}(\omega) \in \mathbb{B}\}.$$

*Remark.* Often times, we write $\Pr(X \in B) = P_X(B)$.

In the context of information theory, we mostly are concerned with real-valued random variables only.

**Definition 1.1.9 ▸ Real-Valued Random Variable**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a **real-valued random variable** over the space is a mapping $X: \Omega \to \mathbb{R}$ such that

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

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

Note that the Borel set over $\mathbb{R}$ is just the family of all open intervals.

Clearly, if $X$ is a real-valued random variable, we have $\{\omega \in \Omega : X(\omega) > x\} \in \mathcal{F}$. Moreover, we claim that

$$\{\omega \in \Omega : X(\omega) < x\} = \bigcup_{y<x} \{\omega \in \Omega : X(\omega) \leq y\}.$$

The proof is quite straightforward and is left to the reader as an exercise. By Definition

1.1.2, this means that

$$\{\omega \in \Omega : X(\omega) < x\} \cup \{\omega \in \Omega : X(\omega) > x\} \in \mathcal{F}.$$

Therefore, $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$. This argument justifies the probabilities $\Pr(X < x)$ and $\Pr(X = x)$. We give a special name to the range of a random variable in computer science.

**Definition 1.1.10 ▸ Alphabet**

Let $X$ be a random variable, the range of $X$ is called an **alphabet**, denoted as $\mathscr{X}$.

Recall that we have defined expectations for discrete and continuous random variables in elementary probability theory. In terms of measure theory, the two formulae can be unified as

$$\mathbb{E}[X] = \int_{\Omega} X(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

Note that $\mathbb{E}[X]$ is a real number while $\mathbb{E}[X \mid Y]$ is a **random variable** formed as a function of $Y$. In general, the following result holds:

**Theorem 1.1.11 ▸ Law of Iterated Expectations**

*Let $X$ and $Y$ be random variables, then* $\mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}[X]$.

## 1.2 Markov Chains

**Theorem 1.2.1 ▸ Markov Inequality**

*If $X$ is a non-negative random variable, then* $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$ *for all* $a > 0$.

**Theorem 1.2.2 ▸ Chebyshev's Inequality**

*For any real-valued random variable $X$ with finite variance,*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| > a\sqrt{\operatorname{Var}(X)}\right) \leq \frac{1}{a^2}$$

*for all* $a > 0$.

A convex function is an overestimate of all linear functions whose values are bounded above by it.

**Theorem 1.2.3 ▸ Jensen's Inequality**

*Let $f$ be a convex function and $X$ be a random variable, then $\mathbb{E}[f(x)] \geq f(\mathbb{E}[X])$.*