1 Probability measures

Definition 1.1. Let $(Ω,F)$ be a measurable space. A probability measure on $(Ω,F)$ is a map $P : F → [0,1]$ satisfying the following two axioms:

(i) $P(\emptyset) = 0$ and $P(Ω) = 1$.

(ii) If $(A_n)_{n=1}^∞ ⊂ F$ is such that $A_n ∩ A_m = \emptyset, ∀n ≠ m$, then $P(∪_{n=1}^∞ A_n) = \sum_{n=1}^∞ P(A_n)$.

In particular, if $A, B ∈ F$ are such that $A ∩ B = \emptyset$, then $P(A ∪ B) = P(A) + P(B)$.

The following properties can be further deduced from the above axioms:

(iii) If $(A_n)_{n=1}^∞ ⊂ F$, then $P(∪_{n=1}^∞ A_n) ≤ \sum_{n=1}^∞ P(A_n)$.

In particular, if $A, B ∈ F$, then $P(A ∪ B) ≤ P(A) + P(B)$.

(iv) If $A, B ∈ F$ and $A ⊂ B$, then $P(A) ≤ P(B)$ and $P(B\setminus A) = P(B) - P(A)$.

In particular, $P(∅) = 1 - P(A)$.

(v) If $A, B ∈ F$, then $P(A ∪ B) = P(A) + P(B) - P(A ∩ B)$.

(vi) If $(A_n)_{n=1}^∞ ⊂ F$ is such that $A_n ⊂ A_{n+1}, ∀n$, then $P(∪_{n=1}^∞ A_n) = \lim_{n→∞} P(A_n)$.

(vii) If $(A_n)_{n=1}^∞ ⊂ F$ is such that $A_n ⊃ A_{n+1}, ∀n$, then $P(∩_{n=1}^∞ A_n) = \lim_{n→∞} P(A_n)$.

Terminology. The triple $(Ω,F,P)$ is called a probability space. Property (ii) is referred to as the σ-additivity (or simply additivity in the finite case) of probability measures.

Example. Let $Ω = \{1,\ldots,6\}$ and $F = P(Ω)$ be the measurable space associated to a die roll. The probability measure associated to a balanced die is defined as

$P_1(\{i\}) = \frac{1}{6}, \ ∀i ∈ \{1,\ldots,6\}$,

and is extended by additivity to all subsets of $Ω$. E.g.,

$P_1(\{1,3,5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.

The probability measure associated to a loaded die is defined as

$P_2(\{6\}) = 1 \ and \ P_2(\{i\}) = 0, \ ∀i ∈ \{1,\ldots,5\}$,

and is extended by additivity to all subsets of $Ω$.

Example. Let $Ω = [0,1]$ and $F = B([0,1])$. One defines the following probability measure on the subintervals of $[0,1]$:

$P([a,b]) = b - a$.

Fact. Carathéodory’s extension theorem states that $P$ can be extended uniquely by σ-additivity to all Borel subsets of $[0,1]$. It is called the Lebesgue measure on $[0,1]$ and is sometimes denoted as $P(B) = |B|$.

Example. Let $Ω = \mathbb{R}$ and $F = B(\mathbb{R})$. One defines the following probability measure on open intervals:

$P([a,b]) = \int_a^b dx \frac{1}{\sqrt{2π}} \exp(-x^2/2)$.

Such a measure can again be uniquely extended to all Borel subsets of $\mathbb{R}$; it is called the (normalized) Gaussian measure on $\mathbb{R}$.

1
**Remarks.** - One can also define the following measure on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\), by setting on open intervals:

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

This measure can be again uniquely extended to all Borel subsets of \(\mathbb{R}\). It is however not a probability measure, as with this definition, one sees easily (using the above axioms) that

\[
P(\mathbb{R}) = \lim_{n \to \infty} P([n, +n[) = \lim_{n \to \infty} 2n = +\infty
\]

This measure is called the **Lebesgue measure** on \(\mathbb{R}\), and is again denoted as \(P(B) = |B|\) for \(B \in \mathcal{B}(\mathbb{R})\).

- We see here that defining first \(P\) on the singletons \(\{x\}\) instead of the open intervals \([a,b[\) would not be a good idea, as we would have \(P(\{x\}) = 0, \ \forall x \in \mathbb{R}\) for both the Gaussian measure and the Lebesgue measure on \(\mathbb{R}\), although these are clearly different.

## 2 Distribution of a random variable

**Definition 2.1.** Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \(X\) be a random variable defined on this probability space. The **distribution** of \(X\) is the map \(\mu_X : \mathcal{B}(\mathbb{R}) \to [0, 1]\) defined as

\[
\mu_X(B) = \mathbb{P}(\{X \in B\}), \quad B \in \mathcal{B}(\mathbb{R}).
\]

**Remark.** The triple \((\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)\) forms a new probability space.

**Notation.** If a random variable \(X\) has distribution \(\mu\), this is denoted as \(X \sim \mu\). Likewise, if two random variables \(X\) and \(Y\) share the same distribution \(\mu\), then they are are said to be **identically distributed** and this is denoted as \(X \sim Y \sim \mu\).

**Example 2.2.** The probability space describing two independent (and balanced) dice rolls is \(\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}, \mathcal{F} = \mathcal{P}(\Omega)\) and

\[
\mathbb{P}((i, j)) = \frac{1}{36}, \quad \forall (i, j) \in \Omega.
\]

Let \(X_1(i, j) = i\) be the result of the first die, and \(Y(i, j) = i + j\) be the sum of the two dice. Then

\[
\mu_{X_1}(\{i\}) = \mathbb{P}(\{X_1 = i\}) = \mathbb{P}(\{(i, 1), \ldots, (i, 6)\}) = \frac{6}{36} = \frac{1}{6}, \quad \forall i \in \{1, \ldots, 6\},
\]

and

\[
\mu_Y(\{2\}) = \mathbb{P}(\{Y = 2\}) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}, \quad \mu_Y(\{3\}) = \mathbb{P}(\{Y = 3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{1}{18}.
\]

More generally:

\[
\mu_Y(\{i\}) = \frac{6 - |7 - i|}{36}, \quad i \in \{2, \ldots, 12\}.
\]

### 2.1 Cumulative distribution function

**Definition 2.3.** Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \(X\) be a random variable defined on this probability space. The **cumulative distribution function** (or **cdf**) of \(X\) is the map \(F_X : \mathbb{R} \to [0, 1]\) defined as

\[
F_X(t) = \mu_X(]-\infty, t]) = \mathbb{P}(\{X \leq t\}), \quad t \in \mathbb{R}.
\]
Fact. The knowledge of $F_X$ is equivalent to the knowledge of $\mu_X$.

From the properties of probability measures, one deduces easily that the cdf of a random variable satisfies the following properties:

(i) $\lim_{t \to -\infty} F_X(t) = 0$, $\lim_{t \to +\infty} F_X(t) = 1$.

(ii) $F_X$ is non-decreasing, i.e. $F_X(s) \leq F_X(t)$ for all $s < t$.

(iii) $F_X$ is right-continuous on $\mathbb{R}$, i.e. $\lim_{t \downarrow 0} F_X(t + \varepsilon) = F_X(t)$, for all $t \in \mathbb{R}$.

Remarks. - $F_X$ has at most a countable number of jumps on the real line. If $F_X$ has a jump of size $p \in [0, 1]$ at $t \in \mathbb{R}$, this actually means that $P(\{X = t\}) = F_X(t) - \lim_{t \downarrow 0} F_X(t - \varepsilon) = p$.

- Any function $F : \mathbb{R} \to \mathbb{R}$ satisfying the above properties (i), (ii) and (iii) is the cdf of a random variable.

2.2 Two important classes of random variables

Discrete random variables.

Definition 2.4. $X$ is a discrete random variable if it takes values in a countable subset $C$ of $\mathbb{R}$, that is, $X(\omega) \in C$ for every $\omega \in \Omega$ (but one may also relax this condition to $P(\{X \in C\}) = 1$).

The distribution of a discrete random variable is entirely characterized by the numbers $p_x = P(\{X = x\})$, where $x \in C$. Notice that $p_x \geq 0$ for all $x \in C$ and that $\sum_{x \in C} p_x = P(\{X \in C\}) = 1$. Moreover,

$$\mu_X(B) = P(\{X \in B\}) = \sum_{x \in B} p_x, \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

and

$$F_X(t) = P(\{X \leq t\}) = \sum_{x \leq t} p_x, \quad \forall t \in \mathbb{R},$$

is a step function.

Example. A binomial random variable $X$ with parameters $n \geq 1$ and $p \in [0, 1]$ (denoted as $X \sim \text{Bi}(n, p)$) takes values in $\{0, \ldots, n\}$ and is characterized by the numbers

$$p_k = P(\{X = k\}) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in \{0, \ldots, n\},$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ are the binomial coefficients.

Continuous random variables.

Definition 2.5. $X$ is a continuous random variable if $P(\{X \in B\}) = 0$ whenever $B \in \mathcal{B}(\mathbb{R})$ is such that $|B| = 0$ (remember that $|B|$ is the Lebesgue measure of $B$).

In particular, this implies that if $X$ is a continuous random variable, then $P(\{X = x\}) = 0 \ \forall x \in \mathbb{R}$ (as $|\{x\}| = 0 \ \forall x \in \mathbb{R}$).

Fact. If $X$ is a continuous random variable according to the above definition, then there exists a function $p_X : \mathbb{R} \to \mathbb{R}$, called the probability density function (or pdf) of $X$, such that $p_X(x) \geq 0 \ \forall x \in \mathbb{R}$, $\int_{\mathbb{R}} p_X(x) \, dx = 1$ and

$$\mu_X(B) = P(\{X \in B\}) = \int_B p_X(x) \, dx, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Moreover,

$$F_X(t) = P(\{X \leq t\}) = \int_{-\infty}^t p_X(x) \, dx, \quad \forall t \in \mathbb{R},$$

is a differentiable function (whose derivative is $F_X'(t) = p_X(t)$).
Example. A Gaussian random variable $X$ with mean $\mu$ and variance $\sigma^2$ (denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$) takes values in $\mathbb{R}$ and has pdf

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$ 

Remark. One could think that the only existing distributions are either discrete or continuous, or a combination of these. It turns out that life is more complicated than that! Some distributions are neither discrete, nor continuous. A famous example is the distribution whose cdf is the devil’s staircase.

## 3 Independence

The notion of independence is a central notion in probability. It is usually defined for events and random variables in elementary probability courses. Nevertheless, as it will become clear below, the independence between $\sigma$-fields turns out to be the most natural concept (remembering that a $\sigma$-field is related to the amount of information one has on a system).

In the three subsections below, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a generic probability space.

### 3.1 Independence of events

One starts by defining the independence of two events in $\mathcal{F}$.

**Definition 3.1.** Two events $A, B \in \mathcal{F}$ are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

**Notation.** $A \perp \perp B$.

**Proposition 3.2.** If two events $A, B \in \mathcal{F}$ are independent, then it also holds that

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) \mathbb{P}(B^c) \quad \text{and} \quad \mathbb{P}(A^c \cap B) = \mathbb{P}(A^c) \mathbb{P}(B).$$

**Proof.** One shows here the first equality (noticing that the other two can be proved in a similar way):

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A \setminus (A \cap B)) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A) \mathbb{P}(B) = \mathbb{P}(A) (1 - \mathbb{P}(B)) = \mathbb{P}(A) \mathbb{P}(B^c).$$

For a collection of more than 2 events, the property $\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$ does not suffice to guarantee that the same property holds for complements of the events $A_i$. A slightly more involved definition of independence is therefore required.

**Definition 3.3.** Let $\{A_1, \ldots, A_n\}$ be a collection of events in $\mathcal{F}$. This collection is independent if

$$\mathbb{P}(A_1^c \cap \ldots \cap A_n^c) = \mathbb{P}(A_1^c) \cdots \mathbb{P}(A_n^c)$$

where $A_i^c$ is either $A_i$ or $A_i^c$, $i \in \{1, \ldots, n\}$.

An intuitive reason why complements should be included in the definition of independence is the following. Let us assume that one rolls a balanced die with four faces. Then the events \{the outcome is 1 or 2\} and \{the outcome is even\} are clearly independent; more precisely, the different informations associated with these events are. So the events \{the outcome is 1 or 2\} and \{the outcome is odd\} are also independent. This motivates the extension of the definition of independence to $\sigma$-fields in the next paragraph.

**Fact.** It can be shown that Definition 3.3 is equivalent to saying that

$$\mathbb{P} \left( \bigcap_{i \in I} A_i \right) = \prod_{i \in I} \mathbb{P}(A_i), \quad \forall I \subseteq \{1, \ldots, n\}.$$ 

From the above fact, one deduces that a collection of events might not be independent, even though its events are two-by-two independent.
3.2 Independence of \( \sigma \)-fields

**Definition 3.4.** Let \( \{ \mathcal{G}_1, \ldots, \mathcal{G}_n \} \) be a collection of sub-\( \sigma \)-fields of \( \mathcal{F} \). This collection is independent if

\[
P(A_1 \cap \ldots \cap A_n) = \prod P(A_i), \quad \forall A_i \in \mathcal{G}_1, \ldots, A_n \in \mathcal{G}_n.
\]

**Example.** Let again \( \{ A_1, \ldots, A_n \} \) be a collection of events in \( \mathcal{F} \). Then the collection of events \( \{ A_1, \ldots, A_n \} \) is independent (according to Definition 3.3) if and only if the collection of \( \sigma \)-fields \( \{ \sigma(A_1), \ldots, \sigma(A_n) \} \) is independent (according to Definition 3.4). In order to see this, observe that \( \sigma(A_i) = \{ \emptyset, A_i, A_i^c, \Omega \} \).

3.3 Independence of random variables

**Definition 3.5.** Let \( \{ X_1, \ldots, X_n \} \) be a collection of random variables defined on \( (\Omega, \mathcal{F}, P) \). This collection is independent if the collection of \( \sigma \)-fields \( \{ \sigma(X_1), \ldots, \sigma(X_n) \} \) is independent.

Since \( \sigma(X_i) = \sigma(\{ X_i \in B \}, B \in \mathcal{B}(\mathbb{R})) \), the collection \( \{ X_1, \ldots, X_n \} \) is independent if and only if

\[
P(\{ X_1 \in B_1, \ldots, X_n \in B_n \}) = P(\{ X_1 \in B_1 \}) \cdots P(\{ X_n \in B_n \}), \quad \forall B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}).
\]

But one also knows that \( \sigma(X_i) = \sigma(\{ X_i \leq t \}, t \in \mathbb{R}) \), so it turns out that \( \{ X_1, \ldots, X_n \} \) is independent if and only if

\[
P(\{ X_1 \leq t_1, \ldots, X_n \leq t_n \}) = P(\{ X_1 \leq t_1 \}) \cdots P(\{ X_n \leq t_n \}), \quad \forall t_1, \ldots, t_n \in \mathbb{R}.
\]

For discrete random variables taking values in a countable set \( C \), this reduces to

\[
P(\{ X_1 = x_1, \ldots, X_n = x_n \}) = P(\{ X_1 = x_1 \}) \cdots P(\{ X_n = x_n \}), \quad \forall x_1, \ldots, x_n \in C.
\]

And for jointly continuous random variables with joint pdf \( p_{X_1, \ldots, X_n} \), this reduces to the classical relation

\[
p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n), \quad \forall x_1, \ldots, x_n \in \mathbb{R}.
\]

The advantage of the above theoretical definition involving \( \sigma \)-fields is the following. Assume \( \{ X_1, \ldots, X_n \} \) is a collection of independent random variables and let \( g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R} \) be Borel-measurable functions. Then one directly deduces from the definition (and the fact that \( g_i(X_i) \) is \( \sigma(X_i) \)-measurable) that \( \{ g_1(X_1), \ldots, g_n(X_n) \} \) is also a collection of independent random variables, which might have been cumbersome to check using any of the other “simpler” definition.

**Example.** Let \( (\Omega, \mathcal{F}, P) \) be a generic probability space and let \( X_0(\omega) = c \in \mathbb{R}, \forall \omega \in \Omega \) be a constant random variable. As \( \sigma(X_0) = \mathcal{F}_0 = \{ \emptyset, \Omega \} \), \( X_0 \) is independent of any other random variable defined on \( (\Omega, \mathcal{F}, P) \).

**Example.** Let \( (\Omega, \mathcal{F}, P) \) be the probability space describing two independent dice rolls in Example 2.2 and let \( X_1(i, j) = i \) and \( X_2(i, j) = j \). One verifies below that these two random variables are indeed independent. It was already shown that \( P(\{ X_1 = i \}) = \frac{1}{6}, \forall i \in \{1, \ldots, 6\} \). Likewise, \( P(\{ X_2 = j \}) = \frac{1}{6}, \forall j \in \{1, \ldots, 6\} \) and

\[
P(\{ X_1 = i, X_2 = j \}) = P(\{ (i, j) \}) = \frac{1}{36} = P(\{ X_1 = i \}) P(\{ X_2 = j \}), \quad \forall (i, j) \in \Omega,
\]

so \( X_1 \) and \( X_2 \) are independent.