Cheatsheet Probability and Statistics

Marc-Olivier Jufer
mjufer@ethz.ch

June 6, 2025

1 Mathematical framework

1.1 Probability space

Def. 1.1. The set Ω is called the **sample space**. An element $\omega \in \Omega$ is called an **outcome** or **elementary experiment**.

Ex. 1.1. *Throw of a die :* $\Omega = \{1, 2, 3, 4, 5, 6\}$

Def. 1.2. A **sigma-algebra** is a subset $\mathcal{F} \subset \mathcal{P}(\Omega)$ satisfying the following properties :

P1. $\Omega \in \mathcal{F}$

P2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$: If A is an event, "not A" is also an event.

P3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$: if $A_1, A_2, \ldots$ are events, then "A_1 or A_2 or $\ldots$" is an event

Ex. 1.2. *Examples of sigma-algebras for* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *:*

- $\mathcal{F} = \{\emptyset, \{1, 2, 3, 4, 5, 6\}\}$
- $\mathcal{F} = \mathcal{P}(\Omega)$
- $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}$

Non examples of sigma-algebras for $\Omega = \{1, 2, 3, 4, 5, 6\}$ *:*

- $\mathcal{F} = \{\{1, 2, 3, 4, 5, 6\}\}$ *: **P2** is not satisfied*
- $\mathcal{F} = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \{1\}, \{2, 3, 4, 5, 6\}, \Omega\}$ *: **P3** is not satisfied*

Def. 1.3. Let Ω a sample space and $\mathcal{F}$ a sigma-algebra. A **probability measure** on $(\Omega, \mathcal{F})$ is a map

$$\mathbb{P} : \mathcal{F} \to [0, 1], \quad A \mapsto \mathbb{P}[A]$$

that satisfies the properties

P1. $\mathbb{P}[\Omega] = 1$

P2. (countable additivity) $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ if $A = \bigcup_{i=1}^{\infty} A_i$ (disjoint union)

Int. A probability measure is a map that associates to each event a number in $[0, 1]$

Ex. 1.3. *For* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *and* $\mathcal{F} = \mathcal{P}(\Omega)$, *the mapping* $\mathbb{P} : \mathcal{F} \to [0, 1]$ *defined by*

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \frac{|A|}{6}$$

is a probability measure on $(\Omega, \mathcal{F})$.

Def. 1.4. Let Ω a sample space, $\mathcal{F}$ a sigma-algebra and $\mathbb{P}$ a probability measure. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.

Int. To construct a probabilistic model, we give

- a sample space Ω : all the possible outcomes of the experiment
- a sigma-algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$: the set of events
- a probability measure $\mathbb{P}$: gives a number in $[0, 1]$ to every event

Def. 1.5. Let $\omega \in \Omega$ (a possible outcome). Let A be an event. We say the event A **occurs** (**does not occur**) (for ω) if $\omega \in A$ ($\omega \notin A$).

1.2 Examples of probability spaces

Def. 1.6. Let Ω be a finite sample space. The **Laplace model** on Ω is the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is defined by

$$\forall A \in \mathcal{F} \quad P[A] = \frac{|A|}{|\Omega|}$$

1.3 Properties of Events

Prop 1.1. (Consequences of definition 1.2). Let $\mathcal{F}$ be a sigma-algebra on Ω. We have

P4. $\emptyset \in \mathcal{F}$

P5. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

P6. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

P7. $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$

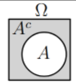
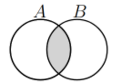
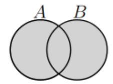
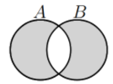
Event	Graphical representation	Probab. interpretation
A^c		A does **not** occur
$A \cap B$		A **and** B occur
$A \cup B$		A **or** B occurs
$A \Delta B$		one and only one of A or B occurs

Figure 1: Representation of set operations

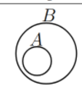
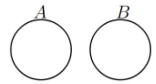
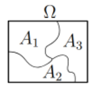
Relation	Graphical representation	Probab. interpretation
$A \subset B$		If A occurs, then B occurs
$A \cap B = \emptyset$		A and B cannot occur at the same time
$\Omega = A_1 \cup A_2 \cup A_3$ with A_1, A_2, A_3 pairwise disjoint		for each outcome ω, one and only one of the events A_1, A_2, A_3 is satisfied.

Figure 2: Representation of set relations

1.4 Properties of probability measures

Prop 1.2. (Consequences of definition 1.3). Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$.

P3. We have $\mathbb{P}[\emptyset] = 0$

P4. (**additivity**) Let $k \geq 1$, let $A_1, \ldots, A_k$ be k pairwise disjoint events, then

$$\mathbb{P}[A_1 \cup \ldots \cup A_k] = \mathbb{P}[A_1] + \ldots + \mathbb{P}[A_k]$$

P5. Let A be an event, then

$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$$

P6. If A and B are two events (not necessarily disjoint), then

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A\ capB]$$

Prop 1.3. (**Monotonicity**). Let $A, B \in \mathcal{F}$, then

$$A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$$

Prop 1.4. (**Union bound**).Let $A_1, A_2, \ldots$ be a sequence of events (not necessarily disjoint), then we have

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

Union bound also applies to a finite collection of events.

Prop 1.5. Let (A_n) be an increasing sequence of events (i.e. $\forall n\ A_n \subset A_{n+1}$). Then

$$\lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\left[\bigcup_{n=1} \infty A_n\right].$$ **increasing limit**

Let (B_n) be a decreasing sequence of events (i.e. $\forall n\ B_n \supset B_{n+1}$). Then

$$\lim_{n\to\infty} \mathbb{P}[B_n] = \mathbb{P}\left[\bigcap_{n=1} \infty B_n\right].$$ **decreasing limit**

[1] i.e. $\Omega = B_1 \cup \ldots \cup B_n$ and the events are pairwise disjoint.

1.5 Conditional probabilities

Def. 1.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Let A, B be two events with $\mathbb{P}[B] > 0$. The **conditional probability of A given B** is defined by

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

Ex. 1.4. *We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ corresponding to the throw of one die. Let $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$. Then*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Prop 1.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Let B be an event with positive probability. Then $\mathbb{P}[\ .\ |B]$ is a probability measure on Ω.

Prop 1.7. (**Formula of total probability**). Let $B_1, \ldots, B_n$ be a partition[1] of the sample space Ω with $\mathbb{P}[B_i] > 0$ for every $i \leq i \leq n$. Then one has

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\, \mathbb{P}[B_i]$$

Prop 1.8. (**Bayes formula**). Let $B_1, \ldots, B_n \in \mathcal{F}$ be a partition of Ω with $\mathbb{P}[B_i] > 0\ \forall i$. For every event A with $\mathbb{P}[A] > 0$ we have

$$\forall i = 1, \ldots, n \quad \mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i]\, \mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A|B_j]\, \mathbb{P}[B_j]}$$

Ex. 1.5. *Test to detect a disease which concerns about* $1/10000$ *of the population. The test gives the right answer* 99% *of the time. If a patient has a positive test, what is the probability that he is actually sick ?*
We modeled the situation as $\Omega = \{0,1\} \times \{0,1\}$. $\mathcal{F} = \mathcal{P}(\Omega)$ and an outcome is $\omega = (\omega_1, \omega_2)$, where ω_1 is 1 if the patient is sick and ω_2 is 1 if the test is positive. Let $S = \{(1,0),(1,1)\}$ be the event that the patient is sick and $T = \{(0,1),(1,1)\}$ the event that the test is positive.
From the hypotheses, we have

$$\mathbb{P}[S] = \frac{1}{10000}, \quad \mathbb{P}[T|S] = \frac{99}{100}, \quad \mathbb{P}[T|S^c] = \frac{1}{100}$$

By applying the Bayes formula to the partition $\Omega = S \cup S^c$, we obtain

$$\mathbb{P}[S|T] = \frac{\mathbb{P}[T|S]\, \mathbb{P}[S]}{\mathbb{P}[T|S]\, \mathbb{P}[S] + \mathbb{P}[T|S^c]\, \mathbb{P}[S^c]} \simeq 0.0098$$

1.6 Independence

Def. 1.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and B are said to be **independent** if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B]$$

A is independent of B iff A is independent of B^c.
If $\mathbb{P}[A] \in \{0, 1\}$, then A is independent of every event.
If A is independent with itself (i.e. $\mathbb{P}[A \cap A] = \mathbb{P}[A]^2$), then $\mathbb{P}[A] \in \{0, 1\}$.

Prop 1.9. Let $A, B \in \mathcal{F}$ be two events with $\mathbb{P}[A], \mathbb{P}[B] > 0$. Then the following are equivalent :

i. $\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B]$: A and B are independent

ii. $\mathbb{P}[A|B] = \mathbb{P}[A]$: the occurrence of B has no influence on A

iii. $\mathbb{P}[B|A] = \mathbb{P}[B]$: the occurrence of A has no influence on B

Def. 1.9. Let I be an arbitrary set of indices. A collection of events $(A_i)_{i \in I}$ is said to be **independent** if

$$\forall J \subset I finite \quad \mathbb{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbb{P}[A_j]$$

Int. Three events A, B and C are independent if the following 4 equations are satisfied :

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\, \mathbb{P}[B]$$

$$\mathbb{P}[A \cap C] = \mathbb{P}[A]\, \mathbb{P}[C]$$

$$\mathbb{P}[B \cap C] = \mathbb{P}[B]\, \mathbb{P}[C]$$

$$\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A]\, \mathbb{P}[B]\, \mathbb{P}[C]$$

2 Random variables and distribution functions

2.1 Abstract definition

Def. 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random variable** (r.v.) is a map $X : \Omega \to \mathbb{R}$ s.t.

$$\forall a \in \mathbb{R} \quad \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$$

Ex. 2.1. *We throw a fair die. The sample space is* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *and we consider the Laplace model* $(\Omega, \mathcal{F}, \mathbb{P})$. *Suppose we gamble on the outcome in such a way that our profit is* -1 *if the outcome is* 1, 2 *or* 3 *;* 0 *if the outcome is* 4 *and* 2 *if the outcome is* 5 *or* 6. *Our profit can be represented by the mapping* X *defined by*

$$\forall \omega \in \Omega \quad X(\omega) = \begin{cases} -1 & \text{if } \omega = 1, 2, 3, \\ 0 & \text{if } \omega = 4, \\ 2 & \text{if } \omega = 5, 6 \end{cases}$$

Since $\mathcal{F} = \mathcal{P}(\Omega)$, *we have* $\{\omega : X(\omega) \leq a\} \in \mathcal{F}$ *for every* a. *Therefore,* X *is a random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$.

Def. 2.2. When events are defined in terms of random variable, we omit the dependence in ω. E.g. for $a \leq b$ we write

$$\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\}$$
$$\{a < X \leq b\} = \{\omega \in \Omega : a < X(\omega) < b\}$$
$$\{X \in \mathbb{Z}\} = \{\omega \in \Omega : X(\omega) \in \mathbb{Z}\}$$

When consider the probability of events as above, we omit the brackets

$$\mathbb{P}[X \leq a] = \mathbb{P}[\{X \leq a\}] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}]$$

2.2 Distribution function

Def. 2.3. Let X be a random variable on a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$. The **distribution function of X** is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$\forall a \in \mathbb{R} \quad F_X(a) = \mathbb{P}[X \leq a]$$

Ex. 2.2. *Same example with the die. Let* X *be the random variable defined as above. For* $a \in \mathbb{R}$ *we have*

$$F_X(a) = \begin{cases} 0 & \text{if } a < -1, \\ 1/2 & \text{if } -1 \leq a < 0, \\ 2/3 & \text{if } 0 \leq a < 2, \\ 1 & \text{if } a \geq 2 \end{cases}$$

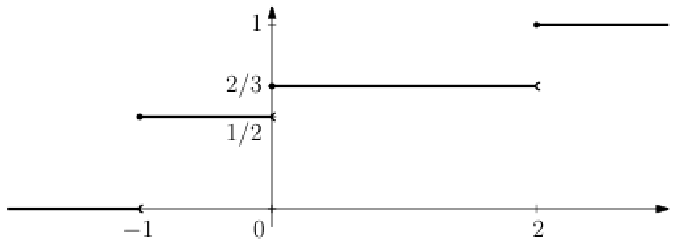


Figure 3: Graph of the distribution function F_X

Prop 2.1. (**Basic identity**). Let $a < b$ be two real numbers. Then

$$\mathbb{P}[a < X \leq b] = F(b) - F(a)$$

Prop 2.2. Let X be a r.v. on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution function $F = F_X : \mathbb{R} \to [0, 1]$ of X satisfies the following properties :

i. F is nondecreasing

ii. F is right continuous[2]

iii. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$

2.3 Independence

Def. 2.4. Let $X_1, \ldots, X_n$ be n random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that they are **independent** if $\forall x_1, \ldots, x_n \in \mathbb{R} \quad \mathbb{P}[X_1 \leq x_1, \ldots, X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \ldots \mathbb{P}[X_n \leq x_n]$.
One can show that $X_1, \ldots, X_n$ are independent iff $\forall I_1 \subset \mathbb{R}, \ldots, I_n \subset \mathbb{R}$ intervals $\{X_1 \in I_1\}, \ldots, \{X_n \in I_n\}$ are independent.

Prop 2.3. (**Grouping**). Let $X_1, \ldots, X_n$ be n independent r.v. Let $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ be some indices and $\phi_1, \ldots, \phi_k$ some functions. Then $Y_1 = \phi_1(X_1, \ldots, X_{i_1}), Y_2 = \phi_2(X_{i_1+1}, \ldots, X_{i_2}), \ldots, Y_k = \phi_k(X_{i_{k-1}+1}, \ldots, X_{i_k})$ are independent.

Def. 2.5. An infinite sequence $X_1, X_2, \ldots$ of random variables is said to be

- **independent** if $X_1, \ldots, X_n$ are independent for every n
- **independent and identically distributed** (i.i.d) if they are independent and they have the same distribution function, i.e. $\forall i, j \quad F_{X_i} = F_{X_j}$.

2.4 Transformation of random variables

We can create r.v. from other r.v. on the same probability space. For example, consider $Z_1 = \exp(X_1), Z_2 = X_1 + X_2$.
Not to forget : r.v. are maps $\Omega \to \mathbb{R}$.
We can work with r.v. as if they were real numbers with the following notation :

Def. 2.6. If X is a r.v. and $\phi : \mathbb{R} \to \mathbb{R}$, then we write

$$\phi(X) := \phi \circ X$$

to to $\phi(X)$ a new mapping $\Omega \to \mathbb{R}$.
We also consider function of several variables. If $X_1, \ldots, X_n$ are n r.v. and $\phi : \mathbb{R}^n \to \mathbb{R}$, then we write

$$\phi(X_1, \ldots, X_n) := \phi \circ (X_1, \ldots, X_n)$$

[2] i.e. $F(a) = \lim_{h \downarrow 0} F(a + h)$ for every $a \in \mathbb{R}$