Probability

Marco Biroli

September 3, 2020

Contents

Chapter 1

Founding Blocks

1.1 Definitions

Definition 1.1.1 (Universe)**.** We consider a random experiment, then the set of all possible outcomes of the experiment is denoted by Ω and is called the universe.

Definition 1.1.2 (Event)**.** An event is usually denoted by E. An event is a set of results for which we can compute the probability.

Definition 1.1.3 (Collection)**.** The collection of all events is denoted by $\mathcal{F}$. Hence $\mathcal{F} \subseteq \mathcal{P}(\Omega)$.

Definition 1.1.4 (Disjoint Events)**.** Two events $A, B \in \mathcal{F}$ are disjoint or incompatible if they cannot occur simultaneously. In other words if $A \cap B = \emptyset$.

Remark. *We require that the collection $\mathcal{F}$ of the events is an algebra of sets.*

Definition 1.1.5 (Algebra of Sets)**.** An element $\mathcal{F}$ is called an algebra of sets if $\mathcal{F} \neq \emptyset$ and:

1. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
2. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

Remark. *For the scope of this course we further require that $\mathcal{F}$ is stable under countable unions. In other words the second condition above is replaced by:*

$$(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$$

Definition 1.1.6 (σ-algebra)**.** A σ-algebra is an algebra of sets where the second condition is replaced by the stronger condition requiring stability under countable union.

Definition 1.1.7 (Probability)**.** The probability $P(E)$ of E is the theoretical value for the proportion of experiments in which E occurs. Thus the probability is a function from $\mathcal{F}$ to $[0, 1]$. Such that:

1. $P(\Omega) = 1$.
2. $A, B \in \mathcal{F}, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.

In other words, P is an additive set function from $\mathcal{F}$ to $[0, 1]$.

Remark. *This definition however is not very well suited to infinite event sets. Then modern probability theory adds a condition to the above.*

Definition 1.1.8 (Modern Probability)**.** A modern probability $P(E)$ of E is a probability with the stronger condition:

$$\forall (A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}, (\forall n, m \in \mathbb{N}, n \neq m \Rightarrow A_n \cap A_m = \emptyset) \Rightarrow P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n)$$

Definition 1.1.9 (Probability Space)**.** A probability space is a triple $(\Omega, \mathcal{F}, P)$. Where Ω is the universe of all possible results, $\mathcal{F}$ is a σ-field on Ω, and P is a modern probability function on $\mathcal{F}$.

Remark. *The mathematical framework which defines probability theory actually comes from another mathematical framework called measure theory. This is why the elements of the σ-field are sometimes called the measurable sets and the probability function is sometimes called a probability measure.*

Definition 1.1.10 (Finite Space)**.** We consider the case where Ω is a finite set, we write $\Omega = \{x_1, \ldots, x_n\}$. The natural σ-field on Ω is $\mathcal{P}(\Omega)$. It is the only σ-field which contains the singletons. Then let P be a probability on Ω and let us set $\forall i \in [\![1, n]\!], p_i = P(\{x_i\})$. Then the numbers p_i satisfy:

$$(\forall i \in [\![1, n]\!], 0 \leq p_i \leq 1) \wedge \sum_{i=1}^{n} p_i = 1$$

Then for any $A \subset \Omega$ we have by additivity that:

$$P(A) = \sum_{x \in A} P(\{x\}) = \sum_{i : x_i \in A} p_i$$

Hence P is completely determined by the numbers p_i.

Remark. *Notice that conversely if we are given the numbers p_i summing to 1 we can define a probability P on Ω by stating $P(\{x_i\}) = p_i$ and P will indeed be a probability measure.*

Definition 1.1.11 (Countable Spaces)**.** We suppose that Ω is countable and we set $\Omega = \{x_n, n \in \mathbb{N}\}$. The natural σ-field on Ω is again the power set of Ω. Then the definitions are an immediate generalization of the ones for a finite space.

Definition 1.1.12 (Continuous Spaces)**.** If we take the simplest example of $\Omega = \mathbb{R}$ then the intuitive σ-field being the power set turns out to be too complicated to be useful. Hence we take for $\mathcal{F}$ the Borel tribe of $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$. The Borel σ-field corresponds to taking a countable union of all possible closed intervals of $\mathbb{R}$.

Definition 1.1.13 (Random Variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable X on $(\Omega, \mathcal{F}, P)$ is map from Ω to $\mathbb{R}$. Which satisfies:

$$\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

Remark. *This definition is equivalent to: for all interval I of $\mathbb{R}$ we have that $X^{-1}(I) \in \mathcal{F}$.*

Notation. *The event $X^{-1}(I)$ is denoted by $\{X \in I\}$ or even simply $X \in I$. Secondly random variables are denoted by capital letters typically X, Y, U, V and their possible values are denoted by the corresponding lowercase letters.*

Definition 1.1.14 (Law of a random variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be a random variable defined on $\Omega \to \mathbb{R}$. The law of X is the probability measure on $\mathbb{R}$ defined by:

$$\forall B \in \mathcal{B}(\mathbb{R}), P_X(B) = P(X \in B)$$

Proof. Let us check that P_X is indeed a probability measure. We have that:

$$P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1.$$

Furthermore let $(B_n)_{n \in \mathbb{N}} \in \mathcal{B}(\mathbb{R})^{\mathbb{N}}$ be a disjoint sequence of Borel sets. Then:

$$P_X\left(\bigcup_{n \in \mathbb{N}} B_n\right) = P\left(X \in \bigcup_{n \in \mathbb{N}} B_n\right) = P\left(\bigcup_{n \in \mathbb{N}} \{X \in B_n\}\right) = \sum_{n \in \mathbb{N}} P(X \in B_n) = \sum_{n \in \mathbb{N}} P_X(B_n)$$

$\square$

Notation. *The law P_X of X is sometimes called the distribution of X. We furthermore say that two variables X, Y have the same law if $P_X = P_Y$. The object of primary interest for a random variable is its law.*

Definition 1.1.15 (Law)**.** Let f be a non-negative function $\mathbb{R} \to \mathbb{R}^+$ which is integrable and $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$. We define next:

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad P(A) = \int_A f(x)\mathrm{d}x$$

This formula defines a probability measure on $\mathbb{R}$, called the probability measure with density function f.

Definition 1.1.16 (Expectation)**.** We say that the random variable X has an expectation or that it is integrable if:

$$\int_{\mathbb{R}} |x| \mathrm{d}P_X(x) < +\infty$$

Then the expectation is defined as:

$$E(X) = \int_{\mathbb{R}} x \mathrm{d}P_X(x) = \int_{\Omega} X \mathrm{d}P = \int_{\omega \in \Omega} X(\omega) \mathrm{d}P(\omega) = \int_{\mathbb{R}} \mathrm{Id}_{\mathbb{R}} \mathrm{d}P_X$$

From this formula we see that the expectation is completely dependent on the law of the random variable.