A primer on measure-theoretic probability

Benedikt Petko
Email: benedikt.petko15@imperial.ac.uk

1 Preliminaries

Definition 1 (Measurable spaces)**.** Let Ω be a set and $\mathcal{F}$ a system of subsets of Ω. The pair $(\Omega, \mathcal{F})$ is said to be a measurable space if

1. $\emptyset \in \mathcal{F}$
2. $\forall A \in \mathcal{F} : A^c \in \mathcal{F}$
3. $\forall (A_i)_{i \in \mathbb{N}}$ countable collections of sets such that $A_i \in \mathcal{F}$ it holds $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$

Remark 2. In the context of probability, the set Ω is called the sample space and $\mathcal{F}$ is called the set of events.

Definition 3 (Probability spaces)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is said to be a probability measure if

1. $\mathbb{P}(\Omega) = 1$
2. $\forall (A_i)_{i \in \mathbb{N}}$ countable collections of pairwise disjoint events it holds that

$$\mathbb{P}(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i)$$

Lemma 4 (Intersections of σ-algebras)**.** Let Ω be a set, I an arbitrary index set and $(\mathcal{F}_i)_{i \in I}$ a collection of σ-algebras on Ω indexed by I. Then

$$\bigcap_{i \in I} \mathcal{F}_i$$

is again a σ-algebra.

Proof. Check the axioms 1.-3. of a σ-algebra.

1. the empty set $\emptyset$ is contained in every $\mathcal{F}_i$, therefore in their intersection $\bigcap_{i \in I} \mathcal{F}_i$ (this is the very definition of what an intersection is)
2. if $A \in \mathcal{F}_i$ for all $i \in I$ then $A^c \in \mathcal{F}_i$ for all $i \in I$, hence A^c lies in the intersection $\bigcap_{i \in I} \mathcal{F}_i$
3. if $(A_k)_{k \in \mathbb{N}}$ are such that $A_k \in \mathcal{F}_i$ for all $i \in I$ then $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{F}_i$ for all $i \in I$, therefore $\bigcup_{k \in \mathbb{N}} A_k$ lies in the intersection $\bigcap_{i \in I} \mathcal{F}_i$

□

Definition 5 (Generated σ-algebras)**.** Let S be a set and $\mathcal{M}$ a collection of subsets (not necessarily a σ-algebra). The σ-algebra generated by $\mathcal{M}$ is defined to be the smallest σ-algebra that contains $\mathcal{M}$, that is

$$\sigma(\mathcal{M}) = \bigcap \{\mathcal{T} : \mathcal{T} \ \sigma\text{-algebra with } \mathcal{M} \subset \mathcal{T}\}$$

Remark 6. According to Lemma 4, the above definition indeed yields a valid σ-algebra.

There is a canonical σ-algebra for Euclidean spaces:

Definition 7 (Borel σ-algebra on $\mathbb{R}$)**.** The Borel σ algebra on $\mathbb{R}$ is defined as

$$\mathcal{B}(\mathbb{R}) := \sigma((a, b) : a, b \in \mathbb{R})$$

The sets in $\mathcal{B}(\mathbb{R})$ are called Borel-measurable sets.

Definition 8 (Preimages)**.** Let U, V be sets and $f : U \to V$ a mapping. The preimage of a subset $B \subset V$ is defined as

$$f^{-1}(B) := \{u \in U : f(u) \in B\}$$

Proposition 9 (Properties of preimages)**.** Preimages have the following three properties:

1. $\forall B \subset V : f^{-1}(V \setminus B) = U \setminus f^{-1}(B)$
2. $\forall (B_i)_{i \in I} : f^{-1}(\bigcap_{i \in I} B_i) = \bigcap_{i \in I} f^{-1}(B_i)$ for an arbitrary index set I
3. $\forall (B_i)_{i \in I} : f^{-1}(\bigcup_{i \in I} B_i) = \bigcup_{i \in I} f^{-1}(B_i)$

Proof. We only prove property 3, the rest follows in the same way. By definition of the preimage and the definition of the set union,

$$\begin{aligned} f^{-1}(\bigcup_{i \in I} B_i) &= \{u \in U : f(u) \in \bigcup_{i \in I} B_i\} \\ &= \{u \in U : \exists i \in I \text{ s. t. } u \in B_i\} \\ &= \bigcup_{i \in I} \{u \in U : f(u) \in B_i\} \\ &= \bigcup_{i \in I} f^{-1}(B_i) \end{aligned}$$

□

Definition 10 (Random variables)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A mapping $X : \Omega \to \mathbb{R}$ is said to be a random variable if

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

Remark 11. In the language of measure theory, the definition says that X is a random variable if it is a measurable map from $\mathcal{F}$ to $\mathcal{B}(\mathbb{R})$

Definition 12 (Distribution of a random variable)**.** Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution of X is a mapping $p_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ defined for any $A \in \mathcal{B}(\mathbb{R})$ as

$$p_X(A) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

Remark 13. The expression on the right is often seen in its shorthand notation as $\mathbb{P}(X \in A)$.

Proposition 14 (Distributions define probability measures)**.** The distribution of a random variable X defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Proof. We need to check the defining properties of a probability measure. First, we have

$$p_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$$

Second, let $(B_i)_{i \in \mathbb{N}}$ be a countable, pairwise disjoint collection of Borel-measurable sets. Then by property 3 of Proposition 9 and $\mathbb{P}$ assumed to be a probability measure,

$$p_X(\bigcup_{i \in \mathbb{N}} B_i) = \mathbb{P}(X^{-1}(\bigcup_{i \in \mathbb{N}} B_i)) = \mathbb{P}(\bigcup_{i \in \mathbb{N}} X^{-1}(B_i)) = \sum_{i \in \mathbb{N}} \mathbb{P}(X^{-1}(B_i)) = \sum_{i \in \mathbb{N}} p_X(B_i)$$

□

Definition 15 (Probability density)**.** Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution p_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A Borel-measurable function $g : \mathbb{R} \to \mathbb{R}$ is said to be the probability density of X if $\forall A \in \mathcal{B}(\mathbb{R})$:

$$p_X(A) = \int_A g(x)dx$$

2 The Lebesgue integral

We lay out how expectation and variance is defined as integrals over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As a motivation, consider first the following provisional definitions.

If a real-valued random variable X has a density $g(x)$ on $\mathbb{R}$ (i.e. it is a continuous random variable), we may define expectation for any $f : \mathbb{R} \to \mathbb{R}$ Borel-measurable as

$$\mathbb{E}[f(X)] := \int_{\mathbb{R}} f(x)g(x)dx$$

where the integral is in the sense of Riemann. In particular, choosing $f(x) = x$ we obtain

$$\mathbb{E}[f(X)] = \mathbb{E}[X] = \int_{\mathbb{R}} xg(x)dx$$

and choosing $f(x) = (x - \mathbb{E}[X])^2$ we get

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{E}} (x - \mathbb{E}[X])^2 g(x)dx$$

which are both familiar definitions from statistics.

If a real-valued random variable X is discrete and we have the probability mass function, we can analogously define expectation and variance by summation instead of Riemann integration.

There are, however, a few limitations to this approach, with the last point being the most crucial one:

- the expectiation is defined differently for discrete and continuous variable - there is an apparent lack of a unified framework
- we are restricted to real-valued random variables by nature of the Riemann integral
- we went at great length to introduce the structure of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the concept of measurability, but there is no more reference to it in this definition of expectation

The solution is to introduce the probability space back to the picture and integrate on Ω rather than on $\mathbb{R}$. After all, the randomness originates from $(\Omega, \mathcal{F}, \mathbb{P})$ while the values on $\mathbb{R}$ are merely observations of this randomness by means of a random variable.

In all of the following, suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Definition 16 (Simple measurable functions)**.** A function $s : \Omega \to \mathbb{R}$ is said to be a simple measurable function if it is of the form

$$s(\omega) = \sum_{i=0}^{n} \alpha_i \mathbb{1}_{A_i}(\omega)$$

for some $n \in \mathbb{N}, \alpha_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$ for $i = 0, \ldots, n$.

Denote the space of simple measurable functions on Ω as $SF(\Omega)$.

Definition 17 (Integrals of simple functions)**.** The Lebesgue integral of a simple measurable function s over an event $A \in \mathcal{F}$ is defined as

$$\int_A s(\omega) d\mathbb{P}(\omega) := \sum_{i=0}^{n} \alpha_i \mathbb{P}(A \cap A_i)$$

Definition 18 (Integrals of measurable functions)**.** Suppose first that $f : \Omega \to \mathbb{R}$ is a measurable, non-negative function. For any $A \in \mathcal{F}$ define the Lebesgue integral of f as

$$\int_A f(\omega) d\mathbb{P}(\omega) := \sup \left\{ \int_A s(\omega) d\mathbb{P}(\omega) : s \in SF(\Omega), 0 \leqslant s \leqslant f \right\}$$

Note that this integral is allowed to take value of $+\infty$.

Now consider f measurable but not necessarily non-negative. Then there exists the decomposition $f(\omega) = f^+(\omega) + f^-(\omega)$ where $f^+(\omega) := \max\{0, f(\omega)\}$ and $f^-(\omega) := \max\{0, -f(\omega)\}$, both of which can be shown to be measurable. If at least one of the integrals $\int_A f^+(\omega) d\mathbb{P}(\omega)$ or $\int_A f^-(\omega) d\mathbb{P}(\omega)$ is finite then we define

$$\int_A f(\omega) d\mathbb{P}(\omega) := \int_A f^+(\omega) d\mathbb{P}(\omega) - \int_A f^-(\omega) d\mathbb{P}(\omega)$$