1. Introduction

Probability Spaces.

A probability space is a measure space $(\Omega, \mathcal{F}, P)$ with $P(\Omega) = 1$.

The *sample space* Ω can be any set, and it can be thought of as the collection of all possible outcomes of some experiment or all possible states of some system. Elements of Ω are referred to as *elementary outcomes*.

The *σ-algebra* (or *σ-field*) $\mathcal{F} \subseteq 2^\Omega$ satisfies

1) $\mathcal{F}$ is nonempty

2) $E \in \mathcal{F} \Rightarrow E^C \in \mathcal{F}$

3) For any countable collection $\{E_i\}_{i \in I} \subseteq \mathcal{F}$, $\bigcup_{i \in I} E_i \in \mathcal{F}$.

Elements of $\mathcal{F}$ are called *events*, and can be regarded as sets of elementary outcomes about which one can say something meaningful. Before the experiment has occurred (or the observation has been made), a meaningful statement about $E \in \mathcal{F}$ is $P(E)$. Afterward, a meaningful statement is whether or not E occurred.

The *probability measure* $P : \mathcal{F} \to [0, 1]$ satisfies

1) $P(\Omega) = 1$

2) for any countable disjoint collection $\{E_i\}_{i \in I}$, $P\left(\bigcup_{i \in I} E_i\right) = \sum_{i \in I} P(E_i)$.

The interpretation is that $P(A)$ represents the chance that event A occurs (though there is no general consensus about what that actually means).

If p is some property and $A = \{\omega \in \Omega : p(\omega) \text{ is true}\}$ is such that $P(A) = 1$, then we say that p holds *almost surely*, or a.s. for short. This is equivalent to "almost everywhere" in measure theory. Note that it is possible to have an event $E \in \mathcal{F}$ with $E \neq \emptyset$ and $P(E) = 0$. Thus, for instance, there is a distinction between "impossible" and "with probability zero" as discussed in Example 1.3 below.

Example 1.1. Rolling a fair die: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^\Omega$, $P(E) = \frac{|E|}{6}$.

Example 1.2. Flipping a (possibly biased) coin: $\Omega = \{H, T\}$, $\mathcal{F} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$, P satisfies $P(\{H\}) = p$ and $P(\{T\}) = 1 - p$ for some $p \in (0, 1)$.

Example 1.3. Random point in the unit interval: $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}_{[0,1]}$ = Borel Sets, P = Lebesgue measure.
The experiment here is to pick a real number between 0 and 1 uniformly at random. Generally speaking, uniformity corresponds to translation invariance, which is the primary defining property of Lebesgue measure. Observe that each outcome $x \in [0, 1]$ has $P(\{x\}) = 0$, so the experiment must result in the realization of an outcome with probability zero.

Example 1.4. Standard normal distribution: $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}$, $P(E) = \frac{1}{\sqrt{2\pi}} \int_E e^{-\frac{x^2}{2}} dx$.

Example 1.5. Poisson distribution with mean λ: $\Omega = \mathbb{N} \cup \{0\}$, $\mathcal{F} = 2^\Omega$, $P(E) = e^{-\lambda} \sum_{k \in E} \frac{\lambda^k}{k!}$.

Why Measure Theory.

Historically, probability was defined in terms of a finite number of equally likely outcomes (Example 1.1) so that $|\Omega| < \infty$, $\mathcal{F} = 2^{\Omega}$, and $P(E) = \frac{|E|}{|\Omega|}$.

When the sample space is countably infinite (Example 1.5), or finite but the outcomes are not necessarily equally likely (Example 1.2), one can speak of probabilities in terms weighted outcomes by taking a function $p : \Omega \to [0, 1]$ with $\sum_{\omega \in \Omega} p(\omega) = 1$ and setting $P(E) = \sum_{\omega \in E} p(\omega)$.

For most practical purposes, this can be generalized to the case where $\Omega \subseteq \mathbb{R}$ by taking a weighting function $f : \Omega \to [0, \infty)$ with $\int_{\Omega} f(x)dx = 1$ and setting $P(E) = \int_E f(x)dx$ (Examples 1.3 and 1.4), but one must be careful since the integral is not defined for all sets E (e.g. Vitali sets*).

Those who have taken undergraduate probability will recognize p and f as p.m.f.s and p.d.f.s, respectively. In measure theoretic terms, $f = \frac{dP}{dm}$ is the Radon-Nikodym derivative of P with respect to Lebesgue measure, m. Similarly, $p = \frac{dP}{dc}$ where c is counting measure on Ω.

Measure theory provides a unifying framework in which these ideas can be made rigorous, and it enables further extensions to more general sample spaces and probability functions.

Also, note that in the formal axiomatic construction of probability as a measure space with total mass 1, there is absolutely no mention of chance or randomness, so we can use probability without worrying about any philosophical issues.

Random Variables and Expectation.

Given a measurable space $(S, \mathcal{G})$, we define an $(S, \mathcal{G})$*-valued random variable* to be a measurable function $X : \Omega \to S$. In this class, the unqualified term "random variable" will refer to the case $(S, \mathcal{G}) = (\mathbb{R}, \mathcal{B})$.

We typically think of X as an observable, or a measurement to be taken after the experiment has been performed.

An extremely useful example is given by taking any $A \in \mathcal{F}$ and defining the indicator function,

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \in A^C \end{cases}.$$

Note that if $(\Omega, \mathcal{F}, P)$ is a probability space and X is an $(S, \mathcal{G})$-valued random variable, then X induces the pushforward probability measure $\mu = P \circ X^{-1}$ on $(S, \mathcal{G})$. Frequently, we will abuse notation and write $P(X \in B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$ for $\mu(B)$.

X also induces the sub-σ-algebra $\sigma(X) = \{X^{-1}(E) : E \in \mathcal{G}\} \subseteq \mathcal{F}$. If we think of Ω as the possible outcomes of an experiment and X as a measurement to be performed, then $\sigma(X)$ represents the information we can learn from that measurement.

In contrast to other areas of measure theory, in probability we are often interested in various *sub-σ-algebras* $\mathcal{F}_0 \subseteq \mathcal{F}$, which we think of in terms of information content.

For instance, if the experiment is rolling a six-sided die (Example 1.1), then $\mathcal{F}_0 = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$ represents the information concerning the parity of the value rolled.

The *expectation* (or *mean* or *expected value*) of a real-valued random variable X on $(\Omega, \mathcal{F}, P)$ is defined as $E[X] = \int_\Omega X(\omega)dP(\omega)$ whenever the integral is well-defined.

Expectation is generally interpreted as a weighted average which gives the "best guess" for the value of the random quantity X.

We will study random variables and their expectations in greater detail soon. For now, the point is that many familiar objects from undergraduate probability can be rigorously and simply defined using the language of measure theory.

That said, it should be emphasized that probability is not just the study of measure spaces with total mass 1. As useful and necessary as the rigorous measure theoretic foundations are, it is equally important to cultivate a probabilistic way of thinking whereby one conceptualizes problems in terms of coin tossing, card shuffling, particle trajectories, and so forth.

* An example of a subset of $[0,1]$ which has no well-defined Lebesgue measure is given by the following construction:

Define an equivalence relation on $[0,1]$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$.
Using the axiom of choice, let $E \subseteq [0,1]$ consist of exactly one point from each equivalence class.
For $q \in \mathbb{Q}_{[0,1)}$, define $E_q = E + q \pmod 1$. By construction $E_q \bigcap E_r = \emptyset$ for $r \neq q$ and $\bigcup_{q \in \mathbb{Q}_{[0,1)}} E_q = [0,1]$.
Thus, by countable additivity, we must have

$$1 = m\left([0,1)\right) = m\left(\bigcup_{q \in \mathbb{Q}_{[0,1)}} E_q\right) = \sum_{q \in \mathbb{Q}_{[0,1)}} m(E_q).$$

However, Lebesgue measure is translation invariant, so $m(E_q) = m(E)$ for all q.
We see that $m(E_q)$ is not well-defined as $m(E_q) = 0$ implies $1 = 0$ and $m(E_q) > 0$ implies $1 = \infty$.

The existence of non-measurable sets can be proved using slightly weaker assumptions than the axiom of choice (such as the Boolean prime ideal theorem), but it has been shown that the existence of non-measurable sets is not provable in Zermelo-Fraenkel alone.

In three or more dimensions, the Banach-Tarski paradox shows that in ZFC, there is no finitely additive measure defined on all subsets of Euclidean space which is invariant under translation and rotation.
(The paradox is that one can cut a unit ball into five pieces and reassemble them using only rigid motions to obtain two disjoint unit balls.)