1. The need for measure theory.

This introductory section is directed primarily to those readers who have some familiarity with undergraduate-level probability theory, and who may be unclear as to why it is necessary to introduce measure theory and other mathematical difficulties in order to study probability theory in a rigorous manner.

We attempt to illustrate the limitations of undergraduate-level probability theory in two ways: the restrictions on the kinds of random variables it allows, and the question of what sets can have probabilities defined on them.

1.1. Various kinds of random variables.

The reader familiar with undergraduate-level probability will be comfortable with a statement like, "Let X be a random variable which has the **Poisson**(5) distribution." The reader will know that this means that X takes as its value a "random" non-negative integer, such that the integer $k \geq 0$ is chosen with probability $\mathbf{P}(X = k) = e^{-5}5^k/k!$. The expected value of, say, X^2, can then be computed as $\mathbf{E}(X^2) = \sum_{k=0}^{\infty} k^2 e^{-5}5^k/k!$. X is an example of a *discrete random variable*.

Similarly, the reader will be familiar with a statement like, "Let Y be a random variable which has the **Normal**$(0, 1)$ distribution." This means that the probability that Y lies between two real numbers $a < b$ is given by the integral $\mathbf{P}(a \leq Y \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$. (On the other hand, $\mathbf{P}(Y = y) = 0$ for any particular real number y.) The expected value of, say, Y^2, can then be computed as $\mathbf{E}(Y^2) = \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$. Y is an example of an *absolutely continuous random variable*.

But now suppose we introduce a new random variable Z, as follows. We let X and Y be as above, and then flip an (independent) fair coin. If the coin comes up heads we set $Z = X$, while if it comes up tails we set $Z = Y$. In symbols, $\mathbf{P}(Z = X) = \mathbf{P}(Z = Y) = 1/2$. Then what sort of random variable is Z? It is not discrete, since it can take on an uncountable number of different values. But it is not absolutely continuous, since for certain values z (specifically, when z is a non-negative integer) we have $\mathbf{P}(Z = z) > 0$. So how can we study the random variable Z? How could we compute, say, the expected value of Z^2?

The correct response to this question, of course, is that the division of random variables into discrete versus absolutely continuous is artificial. Instead, measure theory allows us to give a common definition of expected value, which applies equally well to discrete random variables (like X above), to continuous random variables (like Y above), to combinations of them (like

Z above), and to other kinds of random variables not yet imagined. These issues are considered in Sections 4, 6, and 12.

1.2. The uniform distribution and non-measurable sets.

In undergraduate-level probability, continuous random variables are often studied in detail. However, a closer examination suggests that perhaps such random variables are not completely understood after all.

To take the simplest case, suppose that X is a random variable which has the uniform distribution on the unit interval $[0,1]$. In symbols, $X \sim$ **Uniform**$[0,1]$. What precisely does this mean?

Well, certainly this means that $\mathbf{P}(0 \le X \le 1) = 1$ It also means that $\mathbf{P}(0 \le X \le 1/2) = 1/2$, that $\mathbf{P}(3/4 \le X \le 7/8) = 1/8$, etc., and in general that $\mathbf{P}(a \le X \le b) = b - a$ whenever $0 \le a \le b \le 1$, with the same formula holding if $\le$ is replaced by $<$. We can write this as

$$\mathbf{P}([a,b]) = \mathbf{P}((a,b]) = \mathbf{P}([a,b)) = \mathbf{P}((a,b)) = b - a\,, \quad 0 \le a \le b \le 1\,. \tag{1.2.1}$$

In words, the probability that X lies in any interval contained in $[0,1]$ is simply the length of the interval. (We include in this the degenerate case when $a = b$, so that $\mathbf{P}(\{a\}) = 0$ for the singleton set $\{a\}$; in words, the probability that X is equal to any particular number a is zero.)

Similarly, this means that

$$\mathbf{P}(1/4 \le X \le 1/2 \ \ \underline{\text{or}} \ \ 2/3 \le X \le 5/6)$$

$$= \ \mathbf{P}(1/4 \le X \le 1/2) \ + \ \mathbf{P}(2/3 \le X \le 5/6) \ = \ 1/4 + 1/6 \ = \ 5/12\,,$$

and in general that if A and B are disjoint subsets of $[0,1]$ (for example, if $A = [1/4,\ 1/2]$ and $B = [2/3,\ 5/6]$), then

$$\mathbf{P}\,(A \cup B) \ = \ \mathbf{P}(A) \ + \ \mathbf{P}(B)\,. \tag{1.2.2}$$

Equation (1.2.2) is called *finite additivity.*

Indeed, to allow for countable operations (such as limits, which are extremely important in probability theory), we would like to extend (1.2.2) to the case of a countably infinite number of disjoint subsets: if $A_1, A_2, A_3, \ldots$ are disjoint subsets of $[0,1]$, then

$$\mathbf{P}\,(A_1 \cup A_2 \cup A_3 \cup \ldots) \ = \ \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) + \ldots\,. \tag{1.2.3}$$

Equation (1.2.3) is called *countable additivity.*

Note that we do *not* extend equation (1.2.3) to *uncountable* additivity. Indeed, if we did, then we would expect that $\mathbf{P}([0,1]) = \sum_{x\in[0,1]} \mathbf{P}(\{x\})$, which is clearly false since the left-hand side equals 1 while the right-hand side equals 0. (There is no contradiction to (1.2.3) since the interval $[0,1]$ is not countable.) It is for this reason that we restrict attention to countable operations. (For a review of countable and uncountable sets, see Subsection A.2. Also, recall that for non-negative uncountable collections $\{r_\alpha\}_{\alpha\in I}$, $\sum_{\alpha\in I} r_\alpha$ is defined to be the supremum of $\sum_{\alpha\in J} r_\alpha$ over finite $J \subseteq I$.)

Similarly, to reflect the fact that X is "uniform" on the interval $[0,1]$, the probability that X lies in some subset should be unaffected by "shifting" (with wrap-around) the subset by a fixed amount. That is, if for each subset $A \subseteq [0,1]$ we define the *r-shift* of A by

$$A \oplus r \equiv \{a+r\,;\, a \in A,\ a+r \leq 1\} \cup \{a+r-1\,;\, a \in A,\ a+r > 1\}, \quad (1.2.4)$$

then we should have

$$\mathbf{P}(A \oplus r) = \mathbf{P}(A)\,, \qquad 0 \leq r \leq 1\,. \quad (1.2.5)$$

So far so good. But now suppose we ask, what is the probability that X is rational? What is the probability that X^n is rational for some positive integer n? What is the probability that X is *algebraic*, i.e. the solution to some polynomial equation with integer coefficients? Can we compute these things? More fundamentally, are all probabilities such as these necessarily even *defined*? That is, does $\mathbf{P}(A)$ (i.e., the probability that X lies in the subset A) even make *sense* for every possible subset $A \subseteq [0,1]$?

It turns out that the answer to this last question is no, as the following proposition shows. The proof requires equivalence relations, but can be skipped if desired since the result is not used elsewhere in this book.

Proposition 1.2.6. *There does not exist a definition of* $\mathbf{P}(A)$*, defined for all subsets* $A \subseteq [0,1]$*, satisfying (1.2.1) and (1.2.3) and (1.2.5).*

Proof (optional). Suppose, to the contrary, that $\mathbf{P}(A)$ could be so defined for each subset $A \subseteq [0,1]$. We will derive a contradiction to this.

Define an equivalence relation (see Subsection A.5) on $[0,1]$ by: $x \sim y$ if and only if the difference $y - x$ is rational. This relation partitions the interval $[0,1]$ into a disjoint union of equivalence classes. Let H be a subset of $[0,1]$ consisting of precisely one element from each equivalence class (such H must exist by the Axiom of Choice, see page 200). For definiteness, assume that $0 \notin H$ (say, if $0 \in H$, then replace it by $1/2$).

Now, since H contains an element of each equivalence class, we see that each point in $(0,1]$ is contained in the union $\bigcup_{\substack{r\in[0,1)\\ r \text{ rational}}} (H \oplus r)$ of shifts of H. Furthermore, since H contains just *one* point from each equivalence class, we see that these sets $H \oplus r$, for rational $r \in [0,1)$, are all disjoint.

But then, by countable additivity (1.2.3), we have

$$\mathbf{P}((0,1]) = \sum_{\substack{r\in[0,1) \\ r \text{ rational}}} \mathbf{P}(H \oplus r)\,.$$

Shift-invariance (1.2.5) implies that $\mathbf{P}(H \oplus r) = \mathbf{P}(H)$, whence

$$1 = \mathbf{P}((0,1]) = \sum_{\substack{r\in[0,1) \\ r \text{ rational}}} \mathbf{P}(H)\,.$$

This leads to the desired contradiction: A countably infinite sum of the same quantity repeated can only equal 0, or ∞, or $-\infty$, but it can never equal 1. ∎

This proposition says that if we want our probabilities to satisfy reasonable* properties, then we *cannot* define them for all possible subsets of $[0,1]$. Rather, we must *restrict* their definition to certain "measurable" sets. This is the motivation for the next section.

Remark. The existence of problematic sets like H above turns out to be *equivalent* to the Axiom of Choice. In particular, we can never define such sets explicitly – only implicitly via the Axiom of Choice as in the above proof.

1.3. Exercises.

Exercise 1.3.1. Suppose that $\Omega = \{1,2\}$, with $\mathbf{P}(\emptyset) = 0$ and $\mathbf{P}\{1,2\} = 1$. Suppose $\mathbf{P}\{1\} = \frac{1}{4}$. Prove that $\mathbf{P}$ is countably additive if and only if $\mathbf{P}\{2\} = \frac{3}{4}$.

Exercise 1.3.2. Suppose $\Omega = \{1,2,3\}$ and $\mathcal{F}$ is the collection of all subsets of Ω. Find (with proof) necessary and sufficient conditions on the real numbers x, y, and z, such that there exists a countably additive probability measure $\mathbf{P}$ on $\mathcal{F}$, with $x = \mathbf{P}\{1,2\}$, $y = \mathbf{P}\{2,3\}$, and $z = \mathbf{P}\{1,3\}$.

Exercise 1.3.3. Suppose that $\Omega = \mathbf{N}$ is the set of positive integers, and $\mathbf{P}$ is defined for all $A \subseteq \Omega$ by $\mathbf{P}(A) = 0$ if A is finite, and $\mathbf{P}(A) = 1$ if A is infinite. Is $\mathbf{P}$ finitely additive?

Exercise 1.3.4. Suppose that $\Omega = \mathbf{N}$, and $\mathbf{P}$ is defined for all $A \subseteq \Omega$ by $\mathbf{P}(A) = |A|$ if A is finite (where $|A|$ is the number of elements in

*In fact, assuming the Continuum Hypothesis, Proposition 1.2.6 continues to hold if we require only (1.2.3) and that $0 < \mathbf{P}([0,1]) < \infty$ and $\mathbf{P}\{x\} = 0$ for all x; see e.g. Billingsley (1995, p. 46).

the subset A), and $\mathbf{P}(A) = \infty$ if A is infinite. This $\mathbf{P}$ is of course not a probability measure (in fact it is *counting measure*), however we can still ask the following. (By convention, $\infty + \infty = \infty$.)
(a) Is $\mathbf{P}$ finitely additive?
(b) Is $\mathbf{P}$ countably additive?

Exercise 1.3.5. **(a)** In what step of the proof of Proposition 1.2.6 was (1.2.1) used?
(b) Give an example of a countably additive set function $\mathbf{P}$, defined on *all* subsets of $[0, 1]$, which satisfies (1.2.3) and (1.2.5), but not (1.2.1).

1.4. Section summary.

In this section, we have discussed why measure theory is necessary to develop a mathematical rigorous theory of probability. We have discussed basic properties of probability measures such as additivity. We have considered the possibility of random variables which are neither absolutely continuous nor discrete, and therefore do not fit easily into undergraduate-level understanding of probability. Finally, we have proved that, for the uniform distribution on $[0, 1]$, it will not be possible to define a probability on every single subset.