Notes on Measure Theory and Measure-theoretic Probability

Contents

I Measure Theory. 2

1 The Riemann Integral. 2

1.1 Review: The Riemann Integral. 2

II Probability Theory. 4

2 The need for Measure Theory. 4

2.1 The Uniform Distribution and Non-measurable sets. 5

3 Axioms of Probability. 9

Part I

Measure Theory.

1 The Riemann Integral.

1.1 Review: The Riemann Integral.

We begin with a few definitions needed before we can get to the definition of the Riemann integral. Let $\mathbf{R}$ denote the complete ordered field of real numbers.

Definition. 1.1: Partition.

Suppose $a, b \in \mathbf{R}$ with $a < b$. A *partition* of $[a, b]$ is a finite list $x_0, x_1, x_2, \ldots, x_n$, where

$$a = x_0 < x_1 < x_2 < \ldots < x_n = b$$

We use a partition $x_0, x_1, \ldots, x_n$ of $[a, b]$ to think of $[a, b]$ as a union of closed subintervals as follows:

$$[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup \ldots \cup [x_{n-1}, x_n]$$

The next definition introduces a clean notation for the infimum and supremum of the values of a function on some subset of its domain.

Definition. 1.2: Infimum and supremum of a function.

If f is a real-valued function and A is a subset of the domain of f, then

$$\inf_A f = \inf\{f(x) : x \in A\}$$
$$\sup_A f = \sup\{f(x) : x \in A\}$$

The lower and upper Riemann sums, which we now define, approximate the area under the graph of a non-negative function (or more generally, the signed area corresponding to a real-valued function).

Definition. 1.3: Lower and Upper Riemann sums.

Suppose $f : [a, b] \to \mathbf{R}$ is a bounded function and P is a partition $x_0, x_1, \dots, x_n$ of $[a, b]$. The *lower Riemann sum* $L(f, P, [a, b])$ and *the upper Riemann sum* $U(f, P, [a, b])$ are defined by

$$L(f, P, [a, b]) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{[x_{j-1}, x_j]} f$$
$$U(f, P, [a, b]) = \sum_{j=1}^{n} (x_j - x_{j-1}) \sup_{[x_{j-1}, x_j]} f$$

Our intuition suggests that for a partition with only a small gap between consecutive points, the lower Riemann sum should be a bit less than the area under the graph, and the upper Riemann sum should be a bit more than the area under the graph.

The pictures in the next example help convey the idea of these approximations. The base of the jth rectangle has length $(x_j - x_{j-1})$ and has height $\inf_{[x_{j-1}, x_j]} f$ for the lower Riemann sum and height $\sup_{[x_{j-1}, x_j]} f$ for the upper Riemann sum.

Part II

Probability Theory.

2 The need for Measure Theory.

From undergraduate probability, we are familiar with statements like: *Let* X *be a random variable which has the* $Poisson(k; \lambda)$ *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 the 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, we are 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 $\mathbf{P}(a \leq Y \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-y2/2} dy$. (On the other hand, $\mathbf{P}(Y = y) = 0$ for any particular real number y.) The expected value of Y can be calculated as $\mathbf{E}(Y^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy$. 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 values. But, it is not absolutely continuous, since for certain values z (specifically, when Z is a non-negative integer) we have $P(Z = z) > 0$. 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) and other kinds of random variables not yet imagined.

2.1 The Uniform Distribution and Non-measurable sets.

In undergraduate-level probability theory, 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 X is a random variable which has the uniform distribution on the unit interval $[0, 1]$. In symbols, $X \sim \mathit{Uniform}[0, 1]$. What precisely does this mean?

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

$$\mathbf{P}([a, b]) = \mathbf{P}((a, b]) = \mathbf{P}([a, b)) = \mathbf{P}((a, b)) = b - a \tag{1}$$

where $0 \leq a \leq b \leq 1$.

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

$$\begin{aligned}
&\mathbf{P}(\{1/4 \leq X \leq 1/2\} \cup \{2/3 \leq X \leq 5/6\}) \\
=&\mathbf{P}\{1/4 \leq X \leq 1/2\} + \mathbf{P}\{2/3 \leq X \leq 5/6\} \\
=&1/4 + 1/6 \\
=&\frac{5}{12}
\end{aligned}$$

and in general that if A and B are disjoint subsets of $[0, 1]$ (A and B are mutually exclusive events, so $P(A \cap B) = 0$), we have:

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) \tag{2}$$

Equation (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 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

$$\mathbf{P}(A_1 \cup A_2 \cup A_3 \ldots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \mathbf{P}(A_3) + \ldots \tag{3}$$

Equation (3) is called *countable additivity*.

Note that we do *not* extend equation (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 equation (3) since the interval $[0,1]$ is not countable). It is for this reason that we restrict attention to countable operations.

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

$$A \oplus r \equiv \{a + r : a \in A, a + r \leq 1\} \cup \{a + r - 1 : a \in A, a + r > 1\} \tag{4}$$

then we should have

$$\mathbf{P}(A \oplus r) = \mathbf{P}(A) \tag{5}$$

where $0 \leq r \leq 1$.

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*, that is, the solution to some polynomial equation with integer coefficients? Can we compute these things? More fundamentally, are all probabilities such as these ever *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 the 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.

Theorem: 2.1: Pathological sets.

There does not exist a definition of $\mathbf{P}(A)$, defined for all subsets $A \subseteq [0,1]$, satisfying (1), (3) and (5).

Proof. 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 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 $[0,1]$ into a disjoint union of equivalence classes. The equivalence class of x, denoted $[x]$ is given by,

$$\begin{aligned} [x] &:= \{y : y - x = r, r \in \mathbf{Q}\} \\ &= \{x + r : r \in \mathbf{Q}, x + r \leq 1\} \end{aligned}$$

To belabor the point, we'd have many equivalence classes such as,

$$\begin{aligned} \left[\frac{1}{2}\right] &:= \left\{\frac{1}{2} + r : r \in \mathbf{Q}, \frac{1}{2} + r \leq 1\right\} \\ \left[\frac{1}{3}\right] &:= \left\{\frac{1}{3} + r : r \in \mathbf{Q}, \frac{1}{3} + r \leq 1\right\} \\ \left[\frac{1}{\sqrt{2}}\right] &:= \left\{\frac{1}{\sqrt{2}} + r : r \in \mathbf{Q}, \frac{1}{\sqrt{2}} + r \leq 1\right\} \\ &\vdots \end{aligned}$$

Let H be a subset of $[0,1]$ consisting of precisely one element from each equivalence class. For definiteness assume that $0 \notin H$ (say if $0 \in H$, 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 $(0,1]$ is contained in the the union $\cup_{r \in [0,1) \cap \mathbf{Q}} (H \oplus r)$ of shifts of H.

$$(0,1] = \bigcup_{r \in [0,1) \cap \mathbf{Q}} (H \oplus r)$$

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 (3), we have:

$$\begin{aligned}\mathbf{P}((0,1]) &= \mathbf{P}\left(\bigcup_{r\in[0,1)\cap\mathbf{Q}} (H \oplus r)\right) \\ &= \sum_{r\in[0,1)\cap\mathbf{Q}} \mathbf{P}(H \oplus r)\end{aligned}$$

Shift invariance (5) implies that $\mathbf{P}(H \oplus r) = \mathbf{P}(H)$, whence

$$1 = \mathbf{P}((0,1]) = \sum_{r\in[0,1)\cap\mathbf{Q}} \mathbf{P}(H)$$

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. $\square$

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 2.1. The existence of problematic sets like H 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.

Remark 2.2. Banach and Tarskii even proved that given any two bounded subsets A and B of $\mathbf{R}^3$ each with non-empty interior, it is possible to decompose A into a certain finite number n of disjoint pieces $A = \bigcup_{i=1}^n A_i$ and B into the same number n of disjoint pieces $B = \bigcup_{i=1}^n B_i$, in such a way, that, for each i, A_i is Euclid-congruent to B_i. So, we can disassemble A and rebuild it as B.

3 Axioms of Probability.

We begin by presenting the minimal properties, we will need to define a Probability measure. Hopefully, the reader will convince himself (or herself) that the two axioms presented in definition are reasonable. From these two axioms, flows the entire theory. In order to present these axioms, however, we need to introduce the concept of a σ-algebra.

Let Ω be an abstract space, that is with no special structure. Let 2^Ω denote all subsets of Ω, including the empty set denoted by Φ. With $\mathcal{A}$ being a subset of 2^Ω, we consider the following properties:

1. $\Phi \in \mathcal{A}$ and $\omega \in \mathcal{A}$

2. If $A \in \mathcal{A}$, then $A^C \in \mathcal{A}$, where A^C denotes the complement of A.

3. $\mathcal{A}$ is closed under finite unions and finite intersections: that is, if $A_1, \dots, A_n$ are all in $\mathcal{A}$, then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i$ are in $\mathcal{A}$ as well.

4. $\mathcal{A}$ is closed under countable unions and countable intersections: that is, if $A_1, A_2, A_3, \dots$ are all in $\mathcal{A}$, then $\bigcup_{i=1}^\infty A_i$ and $\bigcap_{i=1}^\infty A_i$ are in $\mathcal{A}$ as well.

***Definition.* 3.1: Algebra and σ-algebra.**

$\mathcal{A}$ is an algebra, if it contains the empty set Φ, the set Ω, it is closed under complementation, and under *finite* unions and intersections.

$\mathcal{A}$ is a σ-algebra, if it contains the empty set Φ, the set Ω, it is closed under complementation, and under *countable* unions and intersections.

Any σ-algebra is an algebra, but not all algebras are σ-algebras.

References

[1] Sheldon Axler. *Measure, Integration and Analysis*, 3rd Edition. Springer, 2007.

[2] David Williams. *Probability with Martingales*, 2nd Edition. Springer Undergraduate texts in Mathematics, 2015.

[3] Steven Schreve. *Stochastic Calculus for Finance I and II*, 2nd Edition. Springer Undergraduate texts in Mathematics, 2015.