The Need for Measure Theory (ref: Ch. 1 Rosenthal)

Q. Why is it necessary to introduce measure theory in order to study probability theory in a rigorous manner?

First: Illustrate the limitations of undergrad-level probability in 2 ways:

1) restrictions on the kinds of random variables (RVs) it allows.

2) question of what sets can have probabilities defined on them.

Types of random variables

- Discrete RV

e.g. Let X be a RV which has the Poisson(5) distribution.

This means that X takes as its value a "random" non-negative integer k, chosen with probability

$$P(X=k) = \frac{e^{-5}\,5^k}{k!}, \quad k=0,1,2,\ldots$$

Here $\underset{\text{parameter}}{\lambda} = 5$. Recall that $E[X] = \underset{\text{mean}}{\lambda} = 5$ in this case

- <u>Continuous RV</u>

  e.g. Let Y be a RV which has the Normal $(0,1)$ distribution. $Y \sim N(0,1)$ – aka Standard Normal

This means that the probability that Y lies between 2 $\mathbb{R}$ numbers $a<b$ is given by the integral

$$P(a<Y<b) = \int_a^b \underbrace{\frac{1}{\sqrt{2\pi}} e^{-y^2/2}}_{f_Y(y) = \text{PDF of } Y} dy$$

[Recalling that $P(Y=y) = 0$ for any particular $y \in \mathbb{R}$]

Y is an absolutely continuous RV.

i.e. a RV whose CDF is a continuous function

- Now suppose we introduce a new RV Z as follows:

  Let X & Y be as above.

  Flip a fair coin.

  $\begin{cases} \text{If coin comes up heads, set } Z = X. \\ \text{If coin comes up tails, set } Z = Y. \end{cases}$

Note that $P(Z = X) = P(Z = Y) = \frac{1}{2}$.

Q. What type of RV is Z?

Not discrete – since it can take on an uncountable # of different values

Not absolutely continuous – since for certain values z, we have $P(Z = z) > 0$ (when z is a non-negative integer)

So how can we study RV Z?

How can we compute, say, $E[Z]$?

<u>Answer</u>: the division of random variables into 2 categories (discrete vs. absolutely continuous) is artificial.

→ <u>Measure Theory</u> allows us to give common definitions to expected value, etc.

1 definition applies to discrete RVs (like X), continuous RVs (like Y), & combinations (like Z).

Measure Theory

* Powerful convergence theorems associated with Lebesgue theory of integration lead to more general, more complete, & more elegant results than the Riemann integral allows.

* Lebesgue's definition of the integral broadens the collection of functions for which the integral is defined.

  measurable functions

* Unnecessary distinctions between discrete & continuous distributions disappear in a measure theoretic context.

  think of events that could happen: this is just a collection of events (set)

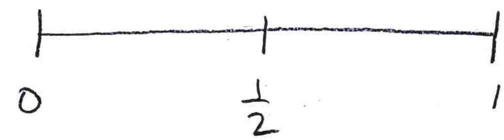
  σ-algebra

The Uniform Distribution & Non-Measurable Sets

Let $X \sim \text{Uniform}(0,1)$. or $X \sim U(0,1)$ - short-hand notation

Means X has the uniform distribution on the interval $[0,1]$.

$$P(0 \leq X \leq 1) = 1$$



$$P(0 \leq X \leq \tfrac{1}{2}) = \tfrac{1}{2}$$

$$P(\tfrac{3}{4} \leq X \leq \tfrac{7}{8}) = \tfrac{1}{8}$$

etc.

In general, we have that for $0 \leq a < b \leq 1$

$$\boxed{P(a \leq X \leq b) = b - a} \quad \leftarrow \text{length of the interval } [a,b] \qquad \text{Eqn (1)}$$

<u>Note</u>: The same formula holds if $\leq$ is replaced by $<$.
Why?

Since X is a continuous RV,
$P(X = x) = 0$ for any $x \in \mathbb{R}$.

$$\underbrace{P([a,b])}_{\text{aka } P(a \le X \le b)} = P((a,b]) = P([a,b)) = P((a,b)) = b-a$$

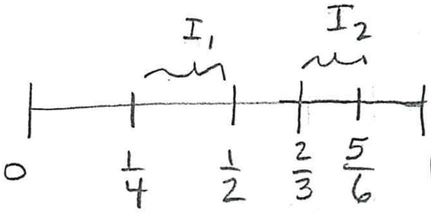
for $0 \le a \le b \le 1$

$$P\left(\tfrac{1}{4} \le X \le \tfrac{1}{2} \text{ OR } \tfrac{2}{3} \le X \le \tfrac{5}{6}\right)$$

$$= P\left(\tfrac{1}{4} \le X \le \tfrac{1}{2}\right) + P\left(\tfrac{2}{3} \le X \le \tfrac{5}{6}\right)$$

$$= \frac{1}{4} + \frac{1}{6}$$

$$= \frac{5}{12}$$



In general, if A & B are disjoint subsets of $[0,1]$,

then $\boxed{P(A \cup B) = P(A) + P(B)}$

Finite additivity property

Extend to a countably infinite # of disjoint subsets:

If $A_1, A_2, \ldots$ are disjoint subsets of $[0,1]$,

then $\boxed{P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots}$ Eqn (2)

$$= \sum_{i=1}^{\infty} P(A_i)$$

Countable additivity

Important Note: We do NOT extend this to uncountable additivity.

If we did, we would expect that

$$P([0,1]) = \sum_{x \in [0,1]} P(\{x\})$$

BUT this is clearly FALSE!

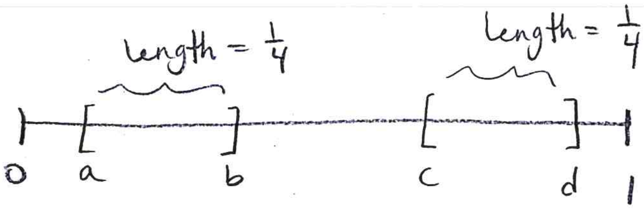
LHS $= 1$

RHS $= 0$ since $P(\{x\}) = P(X = x) = 0$ for any $x \in [0,1]$

Hence, we restrict our attention to countable operations.

Let $\underset{\text{subset}}{A \subseteq [0,1]}$. Then $\underbrace{P(X \in A)}_{\text{Probability that } X \text{ lies in some subset of } [0,1]}$ should be

unaffected by "shifting" (with wrap-around) the subset by a fixed amount.



$c = a + r$

$d = b + r$

shift $[a,b]$ by a constant $r > 0$.

def: We define the r-shift of A ($A \subseteq [0,1]$) by

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

Then, it follows that

$$\boxed{P(A \oplus r) = P(A), \quad 0 \leq r \leq 1.} \qquad \text{Eqn (3)}$$

So far so good.

Now, suppose we ask:

- What is the probability that X is rational?
- " " X^n is rational for some positive integer n?
- What is the probability that X is algebraic, i.e. the solution to some polynomial eqn with integer coefficients?

Q. Can we compute these??

Q. More fundamentally, are all probabilities such as these even defined??

That is, does $P(A) = P(X \in A)$ where $A \subseteq [0,1]$ even make sense for every possible subset A?

Answer: No. (See Prop. below)

Proposition: There does not exist a definition of $P(A)$, defined for all subsets $A \subseteq [0,1]$, that satisfies Equations (1), (2), and (3).

$P([a,b]) = b-a$ r-shift

countable additivity

Proof (sketch): Via Contradiction.

Suppose that $P(A)$ is defined for each subset $A \subseteq [0,1]$. Define an equivalence relation on $[0,1]$ by: $x \sim y$ iff $y - x$ is rational. This relation partitions $[0,1]$ into a disjoint union of equiv. classes.

Let H be a subset of $[0,1]$ consisting of precisely 1 element from each equivalence (H must exist by Axiom of choice). Assume $0 \notin H$; if it is, replace it by $\frac{1}{2}$.

Since H contains an element of each equiv. class, we see that each point in $(0,1]$ is contained in the union $\bigcup_{r\in[0,1)} (H \oplus r)$ of shifts of H.

Furthermore, since H contains just 1 point from each equiv. class, we see that these sets $H \oplus r$, for rational $r \in [0,1)$, are all disjoint.

But then, countable additivity

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

Shift invariance $\Rightarrow$ $P(H \oplus r) = P(H)$

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

This leads to a contradiction: a countably ∞ sum of the same non-neg. quantity can only equal 0 or ∞, but never 1. ▨

Main Point: If we want 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 "measureable" sets.

$\longrightarrow$ Motivation for next section.

Summary

- Discussed why measure theory is necessary to develop mathematical rigorous theory of probability.
- Considered the possibility of RVs that are neither discrete nor (abs.) continuous, don't fit into undergrad. probability.
- Proved that (for $U(0,1)$ distn), it is not possible to define a probability on every single subset.