PROBABILITY: A GRADUATE COURSE

Q: How to incorporate randomness into science/math?

Ans: measure theory.

- Measure $\longrightarrow$ probability
- Measurable sets $\longrightarrow$ events
- Measurable functions $\longrightarrow$ random variables
- Lebesgue integral $\longrightarrow$ expectation
- Radon-Nikodym derivative $\longrightarrow$ conditional expectation

So let's recall the basics of measure theory:

Def (σ-algebra) Let Ω be $\forall$ set.

A collection of subsets $\mathcal{F} \subset 2^{\Omega}$ is called a σ-algebra if:

(i) $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$

(ii) $E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$

i.e. $\mathcal{F}$ is closed under complement and countable union

The sets $E \in \mathcal{F}$ are called measurable; the pair $(\Omega, \mathcal{F})$ is called a measurable space

Remarks ① A σ-algebra is automatically closed under countable intersection:
(DIY: de Morgan's law)

② We automatically have $\Omega \in \mathcal{F}, \ \emptyset \in \mathcal{F}$
$(\Omega = E \cup E^c, \quad \emptyset = \Omega^c)$

Def (measure) Let $(\Omega, \mathcal{F})$ be a measurable space.

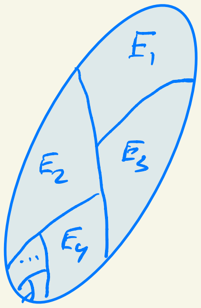
A function $\mu: \mathcal{F} \to \mathbb{R} \cup \{\infty\}$ is called a measure if

(i) $\mu(\emptyset) = 0$

(ii) $\forall E \in \mathcal{F}, \ \mu(E) \geq 0$

(iii) $\forall E_1, E_2, \ldots \in \mathcal{F}$ disjoint, $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$ i.e. "countably additive"

The triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.



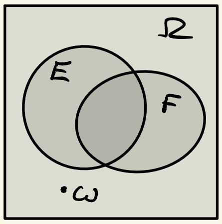
Remark The series in (iii) has nonnegative terms $\Rightarrow$ either converges or diverges to $+\infty$.
It's OK either way.

Q: LHS does not depend on the ordering of E_i. Why does RHS not depend on the ordering?

Teleporting to Probability :

Def Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.
If $\mu(\Omega) = 1$, then μ is called a <u>probability measure</u> or just "probability" and is denoted by $\mathbb{P}$.
In this case, $(\Omega, \mathcal{F}, \mu)$ is called a <u>probability space</u>
Ω is called a <u>sample space</u>
Points $\omega \in \Omega$ are called <u>outcomes</u> (of an "experiment")
Sets $A \in \mathcal{F}$ are called <u>events</u>

Think of an <u>experiment</u>. $\Omega = \{\text{possible outcomes}\}$
$E = \{\text{some outcomes}\} = \text{event}$
$\mathbb{P}(E) = \text{probability of } E$



<u>Set operations</u>:

$E \cap F$ = "E <u>and</u> B occur"
$E \cup F$ = "E <u>or</u> B occur" (or both)
E^c = "E does <u>not</u> occur"
$E \setminus F = E \cap F^c$ = "E but not F occurs"
$E \subset F \iff$ "E <u>implies</u> F"
$E \cap F = \emptyset \iff$ "E and F are mutually exclusive"

EXAMPLES

① <u>Flip a fair coin</u>
$\Omega = \{H, T\}$ $\quad \mathcal{F} = 2^{\Omega} = \{\emptyset, H, T, \{H,T\}\}$
$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$, $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$.

② <u>Flip a fair coin twice</u>
$\Omega = \{H, T, HT, TH, TT\}$
$$\mathcal{F} = 2^{\Omega} = \{\underbrace{\emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH,HT\}, \{HH,TH\}, \ldots}_{2^4 = 16 \text{ elements}}\}$$
$\mathbb{P}(\{HH\}) = \mathbb{P}(\{HT\}) = \mathbb{P}(\{TH\}) = \mathbb{P}(\{TT\}) = \frac{1}{4}$,
and extend to other events by additivity, e.g.
$\mathbb{P}(\text{at least one tail}) = \mathbb{P}(\{HT, TH, TT\}) = \mathbb{P}(\{HT\}) + \mathbb{P}(\{TH\}) + \mathbb{P}(\{TT\}) = \frac{3}{4}$.
$\mathbb{P}$ = "Uniform probability" on Ω.

③ (Waiting time) Flip a coin until the first T

$\Omega = \{T, HT, HHT, HHHT, HHHHT, \dots\}$ ← infinite

$\mathcal{F} = 2^{\Omega}$ ← uncountable

$P(\{T\}) = \frac{1}{2}$, $P(\{HT\}) = \frac{1}{4}$, $P(\{HHT\}) = \frac{1}{8}$, ... (note: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$)

Extend P to $\mathcal{F}$ by countable additivity

(Ex: check that this is well defined)

- P(it takes more than 3 flips to get a tail) $= P(\{HHHT\}, \{HHHHT\}, \{HHHHHT\}, \dots)$

$$= P(\{HHHT\}) + P(\{HHHHT\}) + P(\{HHHHHT\}) + \dots$$

$$= \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^6} + \dots = \frac{1}{2^3} = \frac{1}{8}.$$

- Remark: P is not a uniform measure on Ω.

In fact there is no uniform measure on Ω (or on $\forall$ infinite sample space Ω)

In particular, one can't talk about "a random number uniformly chosen from $\mathbb{R}$"

④ (random permutations)

$\Omega = S_n$ = symmetric group. Elements = permutations of $\{1, \dots, n\}$.

$|\Omega| = n!$ $\quad \mathcal{F} := 2^{\Omega}$

$$P := \text{uniform}: \left\{ \begin{array}{l} \text{make } P(\pi) = \frac{1}{n!} \ \forall \pi \in S_n \\ \text{\& extend by additivity} \end{array} \right\} \Rightarrow P(E) = \frac{|E|}{n!} \quad \forall E \subset S_n$$

- Q: n people form a line at random.

What is the probability that Alice is in front of Bob?

Ans: $\frac{1}{2}$ by symmetry ($n!/2$ permutations with $\pi(a) > \pi(b)$)

(5) <u>(Statistical mechanics)</u> Consider a physical system X

$\Omega :=$ set of possible states X can be in

$E(\omega) :=$ energy of X in state $\omega \in \Omega$ ("Hamiltonian")

Probability that X is in state ω:

$$P(\omega) := \frac{1}{Z} \exp\left(-\frac{E(\omega)}{kT}\right)$$

Here $k =$ Boltzmann constant (absolute)

$T =$ thermodynamic temperature

$Z = \sum_{\omega \in \Omega} \exp\left(-E(\omega)/kT\right) =$ normalizing const ("partition function")

assume Ω is finite

- <u>High temperatures</u>: if $T \to \infty$, $P(\{\omega\}) \to \frac{1}{Z}$ $\forall \omega \in Z$

$\Rightarrow P \to$ uniform probability on Ω

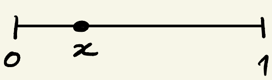
i.e. system is equally likely to be in $\forall$ state

- <u>Low temperatures</u>: if $T \to 0$, $P(\{\omega\}) \to \begin{cases} 1 & \text{if } \omega = \omega_0 \\ 0 & \text{if } \omega \neq \omega_0 \end{cases}$

where $\omega_0 =$ state with lowest energy, i.e. $E(\omega_0) = \min_{\omega \in \Omega} E(\omega)$.

$\Rightarrow$ system freezes in the lowest-energy state.

(6) Break a stick at a random point:



$\Omega = [0,1]$, $P =$ uniform, i.e. Lebesgue measure.

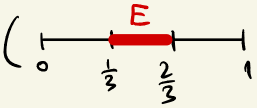
$\mathcal{F} =$ Borel σ-algebra $=$ smallest σ-algebra that contains all open sets.

Q: Why not 2^Ω? $\nexists$ extension of the concept of Lebesgue meas. from intervals (length) to all subsets of $[0,1]$!

<u>Note</u>: $P(\{\omega\}) = 0$ $\forall \omega \in [0,1]$

This is why P assigns probabilities to <u>events</u> $E \in \mathcal{F}$ rather than <u>outcomes</u> $\omega \in \Omega$.

<u>Example</u>: $P\{$neither piece is more than 2× longer than the other$\}$

$= P\Big($  $\Big) = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$.