EECS 126 Lecture 1:

- lab 1 (optional)
- hw 0
- hw 1 (assn next wk)
  ↳ assigned Wed, due Tues
- walrand → problems/applications

Probability : math + way of thinking abt uncertainty → quantification & decision-making/modeling

↳ tools for modifying & quantifying randomness

- inference → inferring "true" meaning/model behind noisy data
- generative models → used by diffussion models
  - ↳ take image, add LOTS of noise to it
  - → reverse this process (run stochastic process in reverse)

Modern Theory of Probability

- axiomatic formulation/starting point
  - axiom: something assumed to be true

Def: Probability space (Ω, F, P) is a triple

Ω : "sample space" = set of samples

F : family of subsets of Ω, called events

↳ $\bigcup_i F_i = \Omega$

P : probability measure

Rules:

1. F is a "σ-algebra" (sigma), contains Ω itself
   - F is closed under complements and countable unions
   - if $A \in F$, then $A^c \in F$
   - if $A_1, A_2 \cdots \in F$ then $\bigcup_{i \geq 1} A_i \in F$

Recall: countable if there's injection from that set to the integers

→ By DeMorgan's, this implies

F is also closed under countable intersections

② P is a fcn that sends events to $[0,1]$

$P: \mathcal{F} \to [0,1]$

↳ P assigns probabilities to events

- Probability measures must obey Kolmogorov Axioms

Kolmogorov Axioms:

I. $P(A) \geq 0 \quad \forall A \in \mathcal{F}$ ← Non-negativity

II. $P(\Omega) = 1$

III. "σ/countable additivity": if $A_1, A_2, \dots \in \mathcal{F}$ is countable disjoint ($A_i \cap A_j = \emptyset$, $i \neq j$) then

$$P\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} P(A_i)$$

Counterexample:

$P = \text{Unif}(0,1)$, then $P(\{x\}) = 0$ if $x \in [0,1]$

By I.

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

contradiction

Examples: Flip a coin w bias p: $\begin{cases} H & \text{w/probability } p \\ T & \text{w/prob } 1-p \end{cases}$

$\Omega = \{H, T\}$

$\mathcal{F} = \underbrace{2^{\Omega}}_{\text{Power set} \Rightarrow \text{all subsets of } \Omega} = \{\emptyset, H, T, \{H,T\}\}$

Probability space for this model

$$\begin{cases} P(H) = p \\ P(T) = 1-p \\ P(\{H,T\}) = 1 \\ P(\emptyset) = 0 \end{cases}$$

Now "n"-independent flips instead of 1 flip

$\Omega = \{H, T\}^n$

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

$P(f_1, \dots, f_n) = p^{\#\{i,\, f_i = H\}} (1-p)^{\#\{i,\, f_i = T\}}$

↳ no "1 right answer" for probability space

Ex:

$\Omega = \{$ all configuration of atoms in the universe $\}$

$A = \{$ set of configuration that lead to flip H $\}$

$B = \{$ " " " " " " " T $\}$

$F = \{ \emptyset, A, B, \Omega = \{A \cup B\} \}$

$P(A) = p$

$P(B) = (1-p)$

$P(\emptyset) = 0$

$P(\Omega) = 1$

↳ difference btwn previous ex is that this is much richer but neither is the wrong way to model

• Probability space often implicitly defined

Consequences of Axioms

↳ usual rules of probability that we're generally familiar with

ex: If $A \in F$ $\quad P(A^c) = 1 - P(A)$ complement

↳ Since F is σ-algebra, then $A^c \in F$ (bc A^c is an event)

$$A \sqcup A^c = \Omega$$

↳ disjoint union

$$1 = \underset{\text{I.}}{P(\Omega)} = P(A \sqcup A^c) \underset{\text{III.}}{=} P(A) + P(A^c)$$

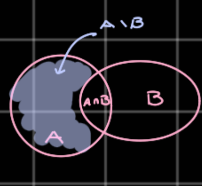
ex2: If $A, B \in F$ $\quad A \subset B \Rightarrow P(A) \leq P(B)$

$$B = A \sqcup \underbrace{(B \setminus A)}_{B \cap A^c}$$

$$P(B) = P(A \sqcup (B \setminus A)) \underset{\text{III.}}{=} P(A) + P(B \setminus A) \underset{\text{I.}}{\geq} P(A)$$

ex3 Inclusion - Exclusion

If $A, B \in F$ $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Proof: $A \cup B,\ A \cap B \in F$

$$A \cup B = B \sqcup (A \setminus B)$$

$$P(A \cup B) \underset{\text{III}}{=} P(B) + P(A \setminus (A \cap B)) \underset{\text{III}}{=} P(B) + P(B) - P(A \cap B)$$

Ex: Probability distribution on the integer (eg. geometric)

Ω = countable set

$F = 2^{\Omega}$, each $\omega \in \Omega$ is an event ($\{\omega\} \in F \ \forall \omega \in \Omega$)

$$\underbrace{P(A) = \sum_{\omega \in A} P(\omega)}_{\text{discrete sample space}} \qquad A \subset \Omega$$

ex * important for conditional prob * Law of Total probability

If $A_1, A_2, \ldots \in F$ partition Ω ($\Omega = \bigsqcup_{i \geq 1} A_i$)

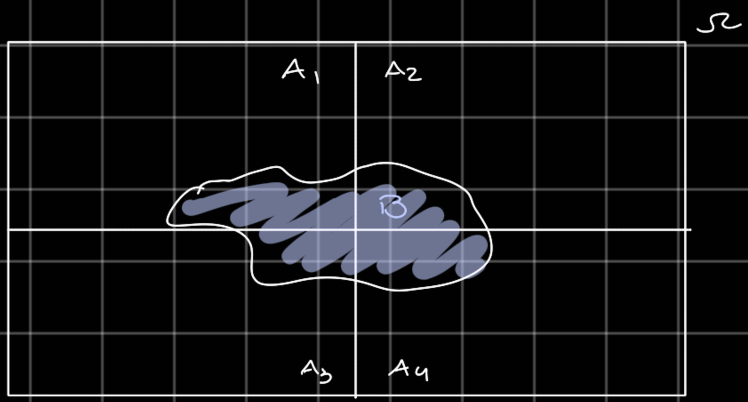
"mutually exclusive, collectively exhaustive"

Then for any $B \in F$:

$$P(B) = \sum_{i \geq 1} P(B \cap A_i)$$

Proof: $B = \bigsqcup_{i \geq 1} (A_i \cap B)$ → apply III

Depiction of Law of Total Prob:



Probabalist / Mathematician: Given (Ω, F, P) what can I say about outcomes / experiments that derive from this model?

opp. perspectives

Statistician perspective: Given outcomes/data, how to choose a good model?

stats $\neq$ probability

<u>Engineer</u> : How do you select a model & use it to draw insight to see how it can be effectively controlled?