Lecture 2

Probability space formalism: (Ω, F, P)

- F: events (a σ-algebra on Ω)
- P: probability measure (axioms)
- Ω: set (of samples)

<u>A I.</u> $A \in F \Rightarrow P(A) \geq 0$ ← negativity

<u>A II.</u> $P(\Omega) = 1$ ← P(anything happening) = 1

<u>A III.</u> $\overbrace{A_1, A_2, \ldots}^{\text{countable}}$ disjoint events ← additivity

$$\Rightarrow P\Big(\bigcup_{i \geq 1} A_i\Big) := \sum_{i \geq 1} P(A_i)$$

Properties (derived last lecture):

- $A \in F \Rightarrow P(A^c) = 1 - P(A)$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ } inclusion-exclusion
- $P(B) = \sum_{i \geq 1} P(B \cap A_i)$ for $A_1, A_2, \ldots \in F$ a partition of Ω
  ↳ (MECE) ↳ total probability rule

Today: Conditional Probability, Independence, Random Vars

<u>Def:</u> If $B \in F$ ("B is an event"), $P(B) > 0$, then we define

$$\text{conditional probability} := \frac{P(A \cap B)}{P(B)} \qquad A \in F$$

↳ intuition: $P(A \mid B)$ = "probability A occurs given we know that B occurred"

↳ more formally: $P(\cdot \mid B)$ gives a restriction of our model (Ω, F, P) to those samples in B

↳ ie, this triple $(B, F|_B, P(\cdot \mid B))$ is a probability space itself, where

$$F|_B = \{ A \cap B : A \in F \}$$

<u>Example</u> (of Cond. Prob):

If $A_1, A_2, \ldots \in F$, $P(A_i) > 0$, (A_i) partition Ω,

↳ we can rewrite the law of total probability in terms of conditional probabilities:

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(B \mid A_i)P(A_i)$$

Bayes Rule

- importance: sometimes it's easy to express $P(\underbrace{B}_{\text{"observation data"}} \mid \underbrace{A}_{\text{"state"}})$

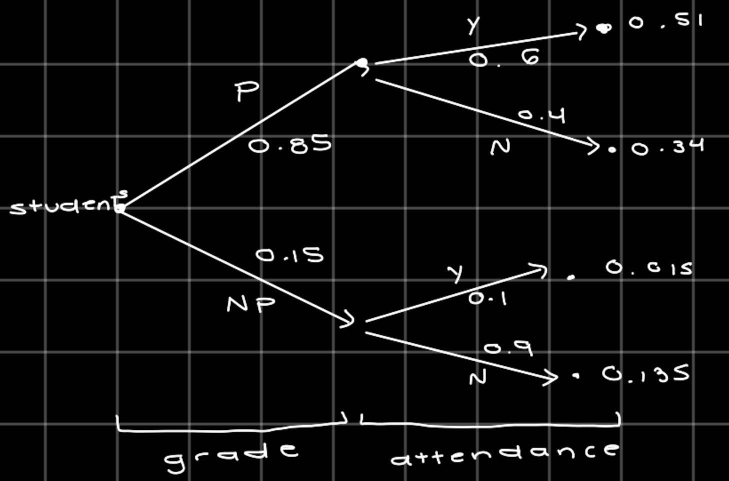
but we're interested in $\underbrace{P(A \mid B)}$

this is the task of inference: how do I get my state from an observation?

$$P(A \mid B) = \frac{P(B \cap A)}{P(B)} = \boxed{\frac{P(A)\,P(B \mid A)}{P(B)}}$$

examples

(a) Suppose we have the following model:



Question:

$$\frac{P(NP \mid N)}{P(NP \mid Y)} = ?$$

$$\frac{P(NP \mid N)}{P(NP \mid Y)} = \frac{P(NP \cap N)}{P(NP \cap Y)} \frac{P(Y)}{P(N)} = \frac{0.135}{0.015} \cdot \frac{0.51}{0.48} = \underset{\sim}{9.9}$$

achievement gap

"how much more likely are students to NP given their attendance?"

9.9x

(b) Dice Rolling: we roll 2 die & the sum is 10.

what is the probability that roll 1 was $= 4$?

$B = \{\text{sum of rolls} = 10\}$

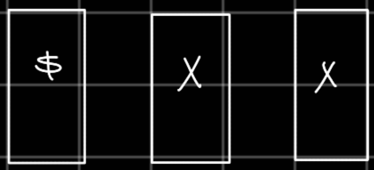
$A = \{\text{first roll} = 4\}$

1st roll, 2nd roll

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{(4,6)\})}{P(\{(4,6),(5,5),(6,4)\})} = \frac{1/36}{3/36} = \frac{1}{3}$$

sample space: all possible tuples of rolls $\rightarrow$ 36 elements

(c) Monty hall problem



1 → Pick door at random 1/3 probability

2 → goat!

3 → switch

- conditioning also gives us a way to usefully decompose intersections of events

Consider: event $A_1 \cap A_2 \cap \cdots \cap A_n$

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(A_1 \mid \bigcap_{i=2}^{n} A_i\right) P\left(\bigcap_{i=2}^{n} A_i\right)$$

$$= P(A_1) P\left(\bigcap_{i=2}^{n} A_i \mid A_1\right)$$

$$= P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 \cap A_2) \cdots P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right)$$

Birthday Paradox

- given n people in a room, what is probability at least 2 share a birthday?

A_i = { person i doesn't share a birthday w/ people $j = 1 \dots i-1$ } ⇝ $365 - (i-1)$ options for bdays

$$P\left(A_i \mid \bigcap_{j=1}^{i-1} A_j\right) = 1 - \frac{(i-1)}{365} = \frac{365 - (i-1)}{365}$$

$$P(\text{No shared b-days}) = P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i \mid \bigcap_{j=1}^{i-1} A_j\right)$$

$$= \prod_{i=1}^{n} \left(1 - \frac{(i-1)}{365}\right)$$

$$\leq e^{-\sum_{i=1}^{n} \frac{i-1}{365}} = e^{-\binom{n}{2}/365}$$

Recall
$1 - x \leq e^{-x}$

$$\Rightarrow P(\text{shared bday}) \geq 1 - e^{-\binom{n}{2}/365}$$

↳ $n = 23$ → P(shared) = 0.5

$n = 261$ → P(shared) = 0.9999...

* as an exercise for yourself, find the probability space
↳ in this problem, it was implicit

Independence

- <u>def</u>: events A, B are independent if $P(A \cap B) = P(A)P(B)$

  ↳ misconception: that disjoint sets are independent

- <u>Note</u>: in special case where

  $P(A) > 0$ & A, B independent $\iff P(B \mid A) = P(B)$

- <u>in general</u>:

  collection $A_1, A_2, \dots \in \mathcal{F}$ are independent events if

  $$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i) \quad \forall \text{ finite sets of indices } S.$$

  subtle
  ↳ this is why we don't talk about only many i.i.d coin flips .... how do you even conceptualize that?

- If $A_1, A_2, \dots \in \mathcal{F}$ are independent then $B_1, B_2, \dots$ are independent where each $B_i = A_i$ or A_i^c. Are the B_i's independent?

  ↳ idea: if you take the complements of events, the complements are also independent

<u>Proof</u>:
$$\bigcap_{i=2}^{n} A_i = \left(\bigcap_{i=1}^{n} A_i\right) \sqcup \left(A_1^c \cap \bigcap_{i=2}^{n} A_i\right)$$

since these are independent, can rewrite as sum of probabilities:

def independence →
$$\prod_{i=2}^{n} P(A_i) = P\left(\bigcap_{i=2}^{n} A_i\right) = P\left(\left(\bigcap_{i=1}^{n} A_i\right) \sqcup \left(A_1^c \cap \bigcap_{i=1}^{n} A_i\right)\right)$$

$$\rightarrow \quad = P\left(\bigcap_{i=1}^{n} A_i\right) + P\left(A_1^c \cap \bigcap_{i=2}^{n} A_i\right)$$

↙ collecting terms & rewriting

$$\underbrace{(1 - P(A_1))}_{P(A_1^c)}\prod_{i=2}^{n} P(A_i) = P\left(A_1^c \cap \bigcap_{i=2}^{n} A_i\right)$$

can't do induction past ∞, must do it for finite sets

$\Rightarrow A_1^c, \dots, A_n$ are independent

↳ by def of independence $\Rightarrow B_1, \dots, B_n$ independent

Conditional independence

- <u>def</u>: If $A, B, C \in \mathcal{F}$ are such that $P(C) > 0$

  and $P(A \cap B \cap C) = P(A \mid C)P(B \mid C)$, then A, B are said to be conditionally independent given C.

- ex: 2 coins with bias p & q

↳ pick coin at random (w.p. $1/2$), flip twice

with probability

$H_i = \{\text{flip } i = H\}$

using law of total probability w/ conditional probabilities

$$P(H_i) = \frac{p+q}{2}$$

$$P(H_1 \cap H_2) = \frac{p^2+q^2}{2} \neq P(H_1)P(H_2) = \left(\frac{p+q}{2}\right)^2$$

$C = \{\text{pick coin } p\}$

↳ then H_1, H_2 conditionally independent given C.