TAM 598 Lecture 2:

PROBABILITY THEORY BASICS

Announcements:

- L3 also online, to be posted on Friday

Part 1: Concepts

① A logical sentence is a statement about the world that can be true or false.

eg) the circumference of the earth at the equator is $40,075 \pm 1$ km

eg) Water is essential for all known forms of life on earth

✱ We will require that the set of logical sentences that we work with be consistent (no contradictions)

We will denote logical sentences with capital letters $A, B, C, \ldots$

We form new logical sentences from old ones using logical connectives:

elbow

- "not A" $\equiv$ "$\neg A$" : the logical sentence that is true when A is false, and false when A is true

- "A and B" $\equiv$ "A, B" $\equiv$ "AB" : the logical sentence that is true when A and B are true, and is otherwise false

- "A or B" $\equiv$ "A + B" : the logical sentence that is true when A or B are true, and otherwise false

inclusive

(2) Probability as a representation of our state of knowledge:

Call I the logical sentence containing ALL information you now have. (Yes, everything. What you learned in school, what your parents taught you, ...)

Let A be a logical sentence, eg, "The result of the next Illinois Fighting Illini basketball game will be an Illinois victory."

→ we do not know if A is true or false. We are uncertain

→ we can use the information I to say something more useful about A.

probability theory is about this;

$$\boxed{p(A|I) \equiv \text{"the probability that } A \text{ is true, given information } I.\text{"}}$$

$p(A|I)$ represents our knowledge about A, given I

(3) But what about frequencies?

- another (probably more common) definition of the probability of an event is the frequency with which it occurs in nature.

- This definition makes sense if the event is something that occurs repeatedly.

- This definition can be restrictive.

  eg) what can we say about an event that can happen only once? What is the probability that life on earth will have ended in a billion years?

  It's hard to apply the frequency interpretation, but we'd still like to use probability to quantify our degree of belief that life on earth will have ended in a billion years

- And, the Bayesian interpretation is compatible with the frequency interpretation in the case of events that occur repeatedly

4. Some "common sense" assumptions, from which <u>ALL</u> of probability theory can be derived:

Consider A is a logical sentence
B is a logical sentence
I is all the information we know

Rules:

- Real numbers represent degrees of plausibility
- Want a qualitative correspondence to common sense
- Want a consistent system:
  - ↳ if there are two ways to calculate a degree of plausibility, they should give the same result
  - ↳ when assigning a probability, we should take into account all evidence
  - ↳ equivalent plausibility assignments must represent equivalent states of knowledge

<u>Cox's Theorem:</u>

these rules are enough to derive the rules of probability theory

$p(A|BI)$ = prob. that A is true, given that B and I are true

called "conditioning"

= prob. that A is true, given that we know B is true

= prob. of A given B

$p(A|BI)$ is a number between 0 and 1

✓ $p(A|B,I) = 1$ when we are certain that A is true if B is true

✓ $p(A|B,I) = 0$ when we are certain that A is false if B is true

✓ $0 < p(A|B,I) < 1$ when we are uncertain about A if B is true

✓ $p(A|B,I) = \frac{1}{2}$ when we are completely ignorant about A if B is true

(5) The Rules of Probability (can be derived from the common sense assumptions)

there are two!

(1) The obvious rule

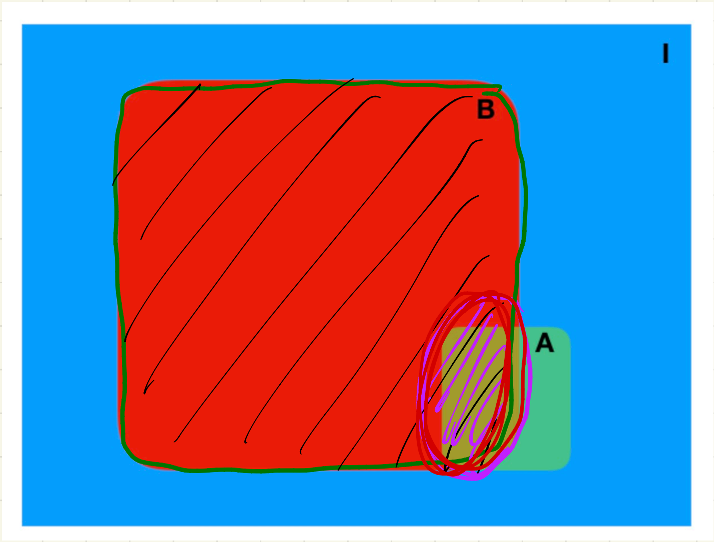
"not A"

$$\boxed{p(A|I) + p(\neg A|I) = 1}$$

"Either A or its negation must be true"

(2) The product rule (aka Bayes' rule, Bayes' Theorem)

$$\underbrace{p(A,B|I)}_{\text{prob. of A and B}} = \underbrace{p(A|B,I)}_{\text{prob. of A given B}} \underbrace{p(B|I)}_{\text{prob. of B}}$$



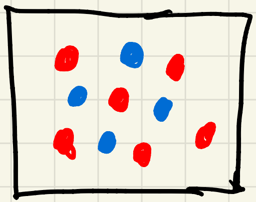
$$p(A,B|I) = \frac{\text{area}_{\text{brown}}}{\text{area}_I}$$

$$p(B|I) = \frac{\text{area}_{\text{red}}}{\text{area}_I}$$

$$p(A|BI) = \frac{\text{area}_{\text{brown}}}{\text{area}_{\text{red}}}$$

$$p(A,B|I) = p(A|BI)\,p(B|I)$$

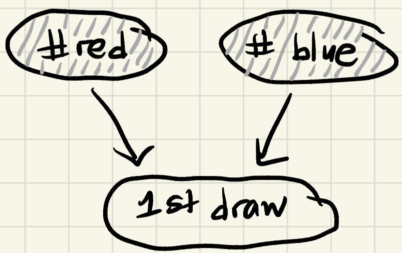
Example: Drawing balls from a box without replacement



Box has 6 red, 4 blue balls

let B_1 = "The first ball we draw is blue."

graphical causal model:



$$p(B_1 | I) = \boxed{\frac{4}{4+6}} = \frac{2}{5}$$

Note: we used the principle of <u>insufficient</u> <u>reason</u>

We can use the obvious rule to find the prob. that the first ball is red

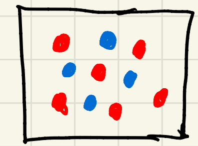
$$p(R_1 | I) = p(\neg B_1 | I)$$

$$= 1 - p(B_1 | I)$$

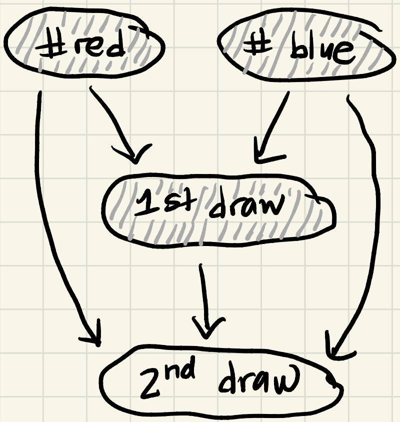
$$= 1 - 2/5$$

$$= 3/5$$

What is the probability that the second draw is red, given that we observe the first draw?



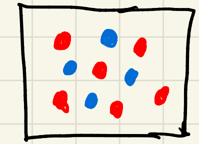
graphical causal model:



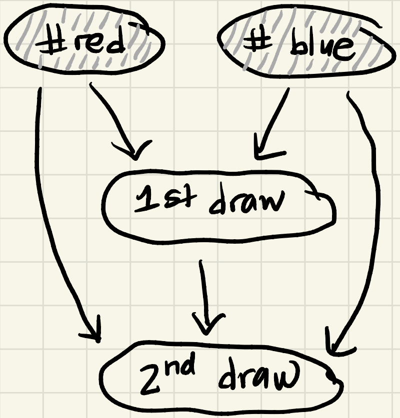
if the first draw is blue : $p(R_2 \mid B_1, I) = \frac{6}{6+3} = \frac{2}{3}$

if the first draw is red : $p(R_2 \mid R_1, I) = \frac{5}{5+4} = \frac{5}{9}$

What is the probability that the first draw is blue and the second draw is red?



graphical causal model:



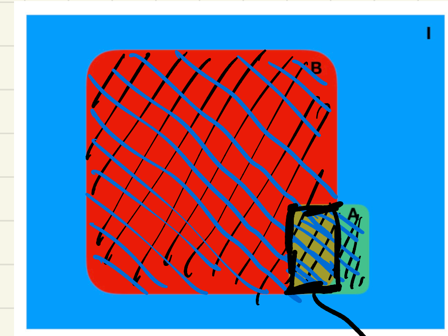
use product rule:

$$p(B_1, R_2 \mid I) = p(R_2 \mid B_1, I)\; p(B_1, I)$$

$$= \left(\frac{2}{3}\right)\left(\frac{2}{5}\right)$$

$$= 4/15$$

Other Rules

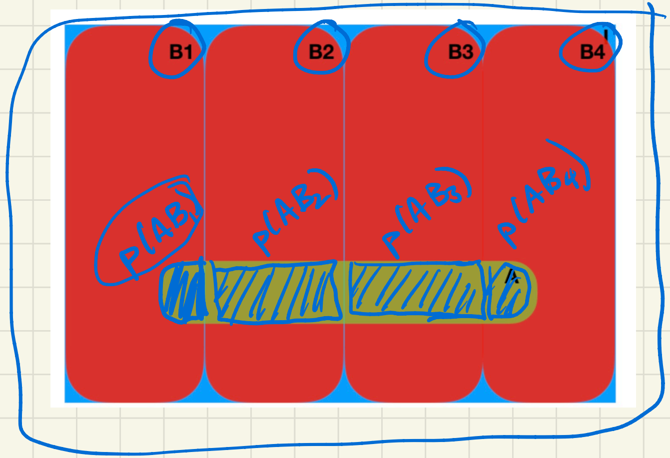


(1) Extension of the Obvious Rule:

inclusive OR

$$\underbrace{p(A+B|I)}_{\text{prob. that A or B is true}} = \underbrace{p(A|I)}_{\text{prob that A is true}} + \underbrace{p(B|I)}_{\text{prob that B is true}} - \underbrace{p(AB|I)}_{\text{prob that A and B are true (double counting correction)}}$$

Other Rules



(2) The sum rule :

Given a sequence of logical sentences $B_1, \dots, B_n$ such that

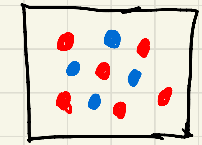
→ one of them is true : $p(B_1 + \dots + B_n \mid I) = 1$

→ they are mutually exclusive : $p(B_i B_j \mid I) = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$

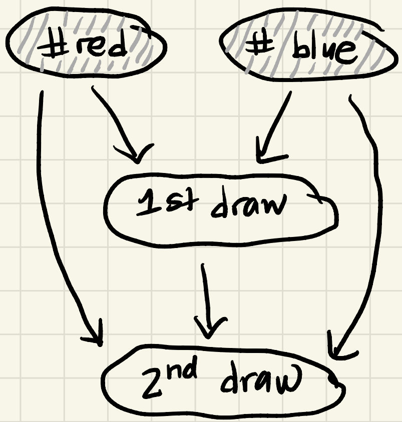
Then for any logical sentence A

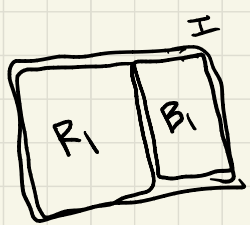
$$p(A \mid I) = \sum_{i=1}^{n} p(AB_i \mid I) = \sum_{i=1}^{n} p(A \mid B_i, I)\, p(B_i \mid I)$$

What is the prob. that the second draw is red?



graphical causal model:





B_1, R_1 cover all possibilities, are mutually exclusive

we use the sum rule:

$$p(R_2) = p(R_2 | B_1)\, p(B_1) + p(R_2 | R_1)\, p(R_1)$$

$$= \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) + \left(\frac{5}{9}\right)\left(\frac{3}{5}\right)$$

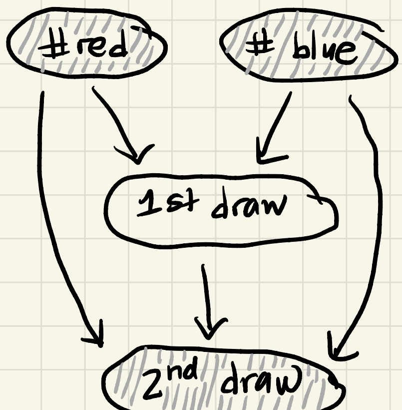
$$= \frac{4}{15} + \frac{1}{3}$$

$$= 3/5$$

Example: so far, our examples have done conditioning by following causal links. But (maybe surprisingly) conditioning does not need to follow causal links.

What is the prob. that the first draw was blue, given that you observe that the second draw is red?

graphical causal model:



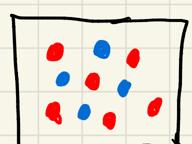
want:

$$p(B_1 \mid R_2, I)$$

product rule:

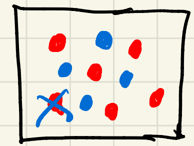
$$p(B_1 \mid R_2, I)\, p(R_2 \mid I) = p(B_1, R_2 \mid I)$$

$$(\quad)\left(\frac{3}{5}\right) = \left(\frac{4}{15}\right)$$



$$p(B_1 \mid R_2, I) = \frac{4}{9} \text{ or } 44\%$$

The probability that the first draw was blue actually increases, given that the second draw was red.



$$p(B_1 \mid R_2, I) = 0.44$$

$$p(B_1 \mid I) = 0.4$$

Is this reasonable? Yes.

- draw a ball w/o seeing it, and you put it in a box
- draw the second ball, and observe that it is red.
- this means you did not pick this red ball in the first draw
- as if we had one less red to worry about on first draw

$$P = \frac{4 \text{ blue}}{4 \text{ blue} + 5 \text{ red}} = \frac{4}{9} \simeq 0.44$$

same result!

Experimenting with "Randomness"

- (*) A coin toss is governed by the laws of classical mechanics, and as such, is deterministic.

- (*) In reality, we cannot perfectly measure or control the variables, so the outcome is random for all practical purposes.

$$X = \begin{cases} T & \text{if } \dfrac{2 v_0 w_0}{g} \bmod (2\pi) \in \left(\dfrac{\pi}{2}, \dfrac{3\pi}{2}\right) \\ H & \text{else} \end{cases}$$

→ V_0 = initial speed $\quad 2.5 \pm 0.2$ m/s

→ w_0 = angular velocity $\quad 200(2\pi) \pm 50$ rad/s

→ g = gravity