Probabilistic and Bayesian Analytics

Andrew W. Moore
School of Computer Science
Carnegie Mellon University

Discrete Random Variables

- A is a Boolean-valued random variable if A denotes an event, and there is some degree of uncertainty as to whether A occurs.
- Examples
- A = The US president in 2023 will be male
- A = You wake up tomorrow with a headache
- A = You have Ebola

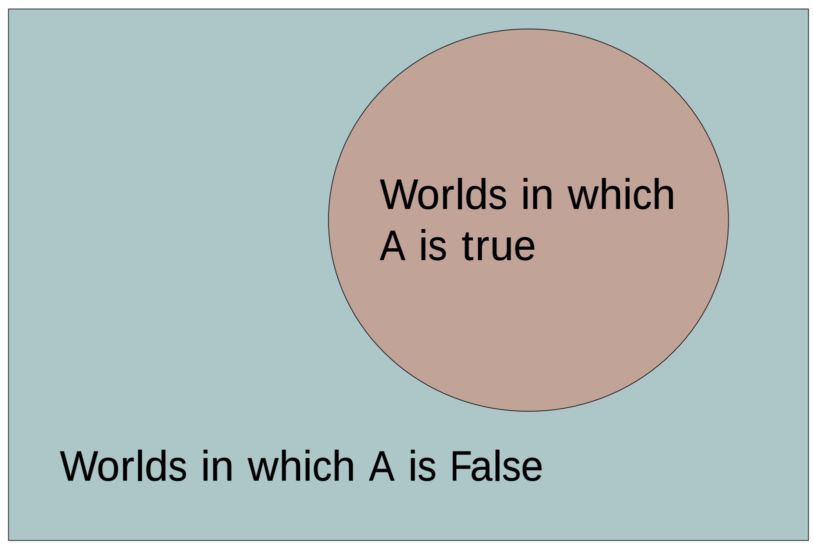
Probabilities

- We write P(A) as “the fraction of possible worlds in which A is true”
- We could at this point spend 2 hours on the philosophy of this.
- But we won’t.

Visualizing A

Event space of
all possible
worlds

Its area is 1



P(A) = Area of
reddish oval

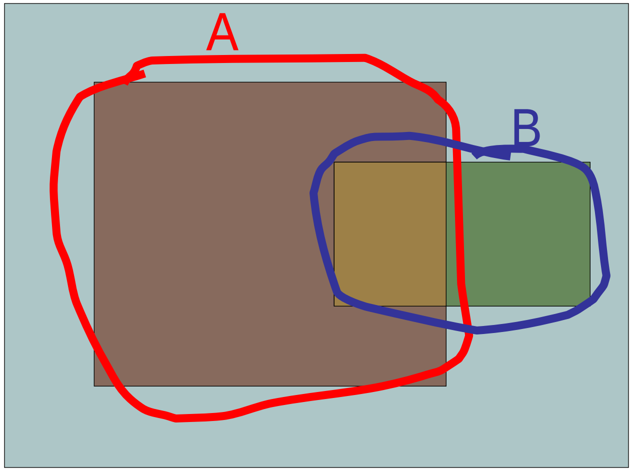

The
Axioms
Of
Probabi
lity

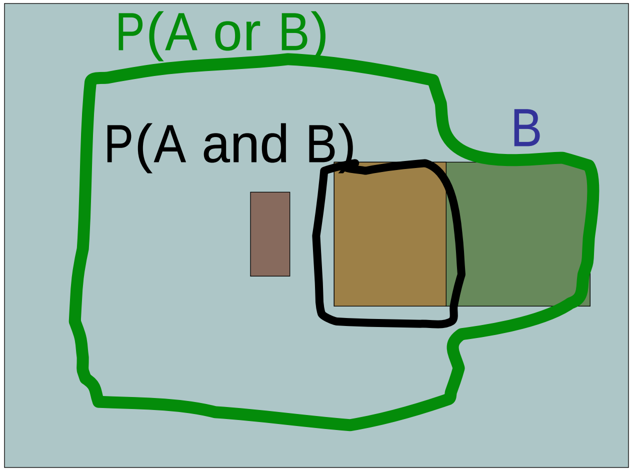
The Axioms of Probability

- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)





Simple addition and subtraction

These Axioms are Not to be Trifled With

- There have been attempts to do different methodologies for uncertainty
  - Fuzzy Logic
  - Three-valued logic
  - Dempster-Shafer
  - Non-monotonic reasoning

- But the axioms of probability are the only system with this property:

  If you gamble using them you can’t be unfairly exploited by an opponent using some other system [di Finetti 1931]

Theorems from the Axioms

- 0 <= P(A) <= 1, P(True) = 1, P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

From these we can prove:

P(not A) = P(~A) = 1-P(A)

- How?

Side Note

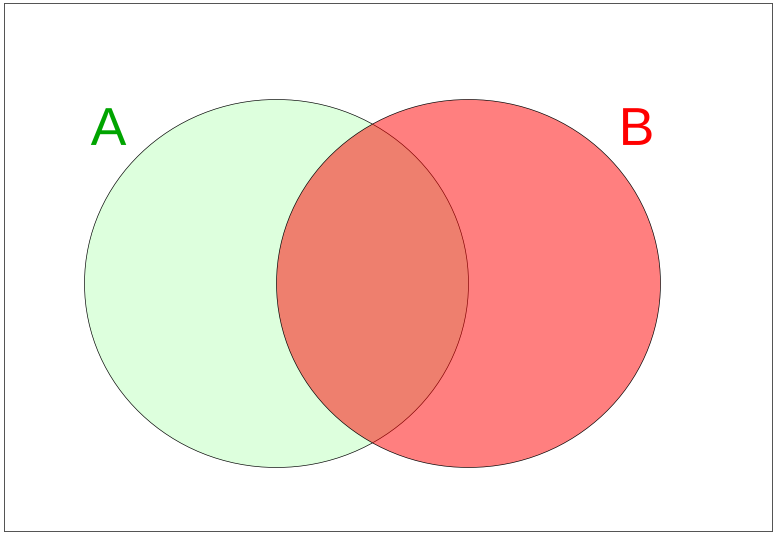
- I am inflicting these proofs on you for two reasons:
  1. These kind of manipulations will need to be second nature to you if you use probabilistic analytics in depth
  2. Suffering is good for you

Another important theorem

- 0 <= P(A) <= 1, P(True) = 1, P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

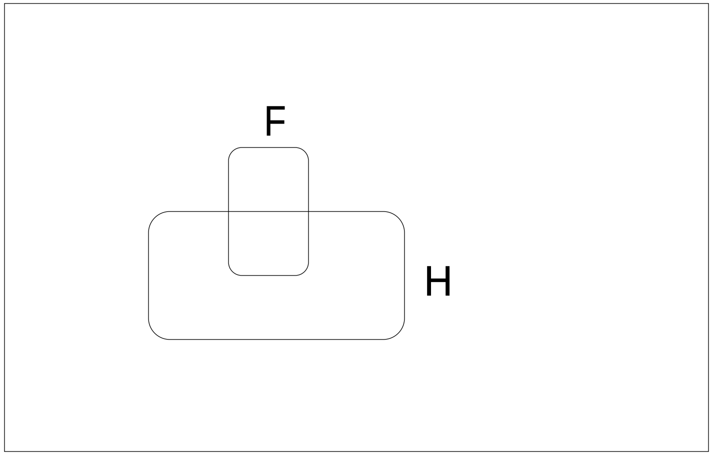
From these we can prove:

P(A) = P(A and B) + P(A and not B)



Conditional Probability

- P(A|B) = Fraction of worlds in which B is true that also have A true

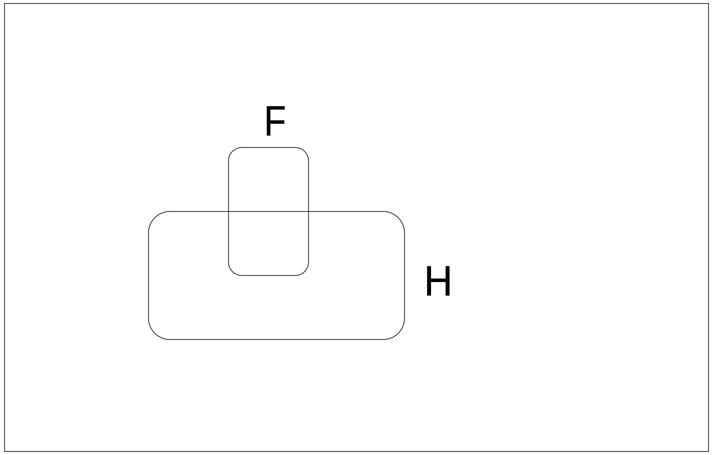


H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

"Headaches are rare and flu is rarer, but if you're coming down with 'flu there's a 50-50 chance you'll have a headache."

Conditional Probability



H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

P(H|F) = Fraction of flu-inflicted worlds in which you have a headache

= #worlds with flu and headache

#worlds with flu

= Area of "H and F" region

Area of "F" region

= P(H ^ F)

P(F)

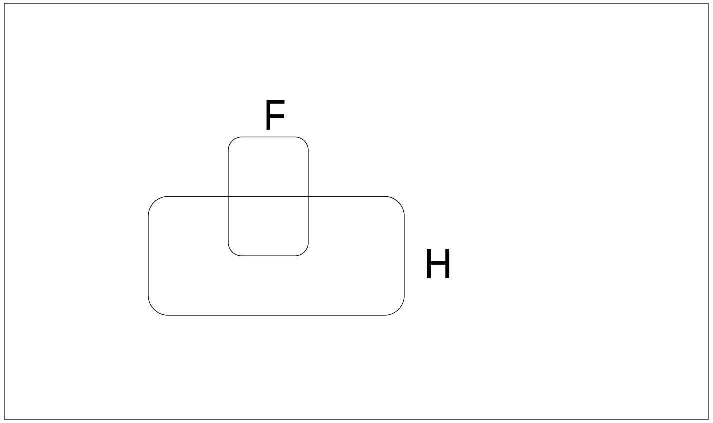
Definition of Conditional Probability

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

Corollary: The Chain Rule

$$P(A \wedge B) = P(A|B) P(B)$$

Probabilistic Inference



H = "Have a headache"
F = "Coming down with Flu"

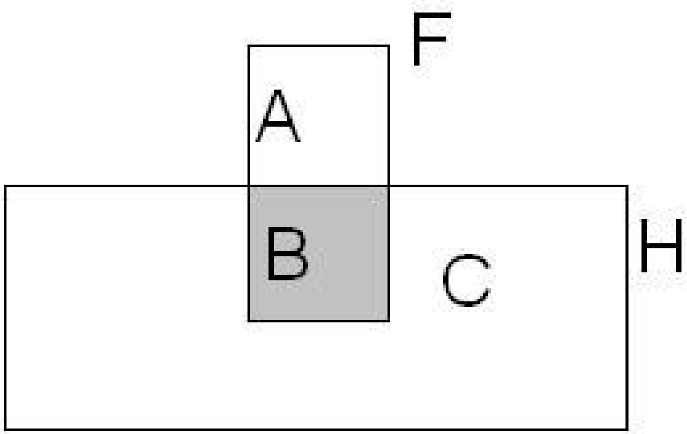
P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

One day you wake up with a headache. You think: "Drat! 50% of flus are associated with headaches so I must have a 50-50 chance of coming down with flu"

Is this reasoning good?

Another way to understand the intuition

Thanks to Jahanzeb Sherwani for contributing this explanation:



Let's say we have P(F), P(H), and P(H|F), like in the example in class.

Areawise, P(F) = A + B,   P(H) = B + C,

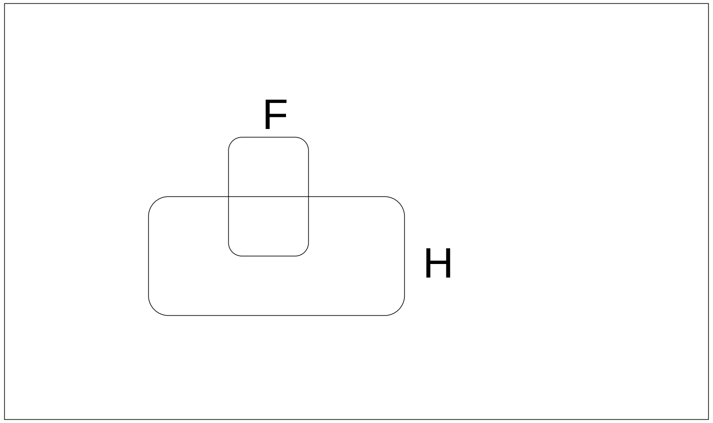
Also, $P(H|F) = \frac{B}{A + B}$

Thus, to get the opposite conditional probability, ie, P(F|H), we need to figure out $\frac{B}{B + C}$

Since we know B / (A+B), we can get B / (B+C) by multiplying by (A+B) and dividing by (B+C). But since we already calculated, A+B = P(F), and B+C = P(H), so we are actually multiplying by P(F) and dividing by P(H). Which is Bayes Rule:

$$P(F|H) = P(H|F) * \frac{P(F)}{P(H)}$$

Probabilistic Inference



H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

$$P(F \text{ and } H) = P(H \mid F) \times P(F) = \frac{1}{2} \times \frac{1}{40} = \frac{1}{80}$$

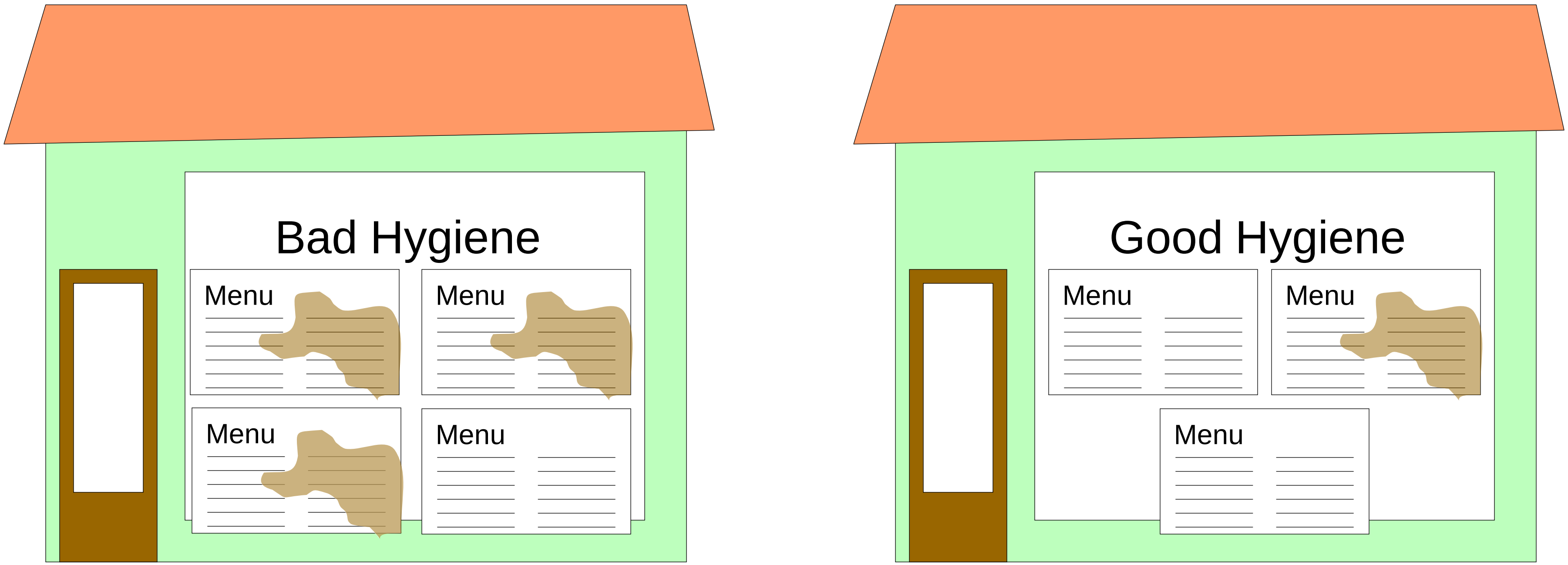
$$P(F \mid H) = \frac{P(F \text{ and } H)}{P(H)} = \frac{1/80}{1/10} = \frac{1}{8}$$

What we just did...

$$P(B|A) = \frac{P(A \wedge B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A)}$$

This is Bayes Rule

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London,* **53:370-418**



- You are a health official, deciding whether to investigate a restaurant
- You lose a dollar if you get it wrong.
- You win a dollar if you get it right
  - Half of all restaurants have bad hygiene
  - In a bad restaurant, ¾ of the menus are smudged
  - In a good restaurant, 1/3 of the menus are smudged
  - You are allowed to see a randomly chosen menu

$$P(B \mid S) = \frac{P(B \text{ and } S)}{P(S)} = \frac{P(S \text{ and } B)}{P(S)}$$

$$= \frac{P(S \text{ and } B)}{P(S \text{ and } B) + P(S \text{ and not } B)}$$

$$= \frac{P(S \mid B)P(B)}{P(S \text{ and } B) + P(S \text{ and not } B)}$$

$$= \frac{P(S \mid B)P(B)}{P(S \mid B)P(B) + P(S \mid \text{not } B)P(\text{not } B)}$$

$$= \frac{\frac{3}{4} \times \frac{1}{2}}{\frac{3}{4} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}} = \frac{9}{13}$$