TEXAS
The University of Texas at Austin

PHIL 313Q: Inductive Logic

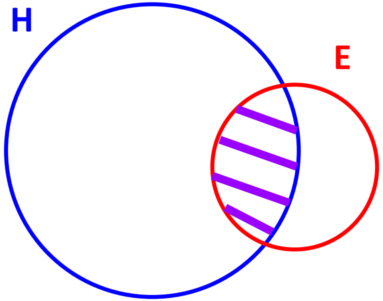
ch. 7 of “Probability and Inductive Logic” by Ian Hacking

Alex Rausch
PhD Candidate, The University of Texas at Austin

TEXAS
The University of Texas at Austin

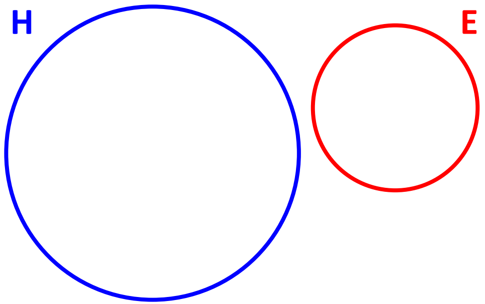
Bayes' Rule

- Suppose you're a scientist. Let **H** be some **H**ypothesis – your scientific theory.
- Let **E** be some piece of **E**vidence – like an observational result of an experiment you're about to run.
- Suppose that before the **E**xperiment, you started with some probability for the **H**ypothesis *on its own*: **Pr(H)** = 50%   (You are 50% confident in **H**.)
- Then you observe the result of the **E**xperiment (**E**).
- **How should you update your confidence in H, given that you observed E?**



Pr(H | E) = Pr(H&E) / Pr(E)

= Pr(H)Pr(E | H) / Pr(E)



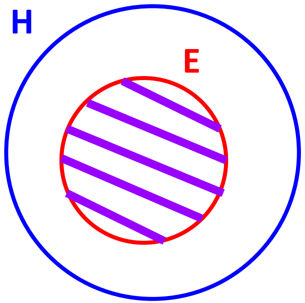
Pr(H | E) = *What, intuitively?*

= Pr(H&E) / Pr(E)

= Pr(H)Pr(E | H) / Pr(E)

= Pr(H) x 0 / Pr(E)

= 0



Pr(H | E) = *What, intuitively?*

= Pr(H&E) / Pr(E)

= Pr(H)Pr(E | H) x (1/Pr(E))

= Pr(H) x [Pr(E)/Pr(H)] x (1/Pr(E))

= 1

Bayes' Theorem

$$Pr(H \mid E) = Pr(H)Pr(E \mid H) / Pr(E)$$

"posterior" "prior" "likelihood" "marginal"

- We know that H and ~H cover the entire space of probability.
  - **Pr(H) + Pr(~H)** = 1
- So **Pr(E)** = **Pr((H & E) v (~H & E))**

  = **Pr(H & E) + Pr(~H & E)**

  = **Pr(H)Pr(E | H)** + **Pr(~H)Pr(E | ~H)**
- Sometimes, the marginal **Pr(E)** might be expanded in the statement of Bayes' Theorem: **Pr(H | E)** = **Pr(H)Pr(E | H)** / [**Pr(H)Pr(E | H) + Pr(~H)Pr(E | ~H)**]

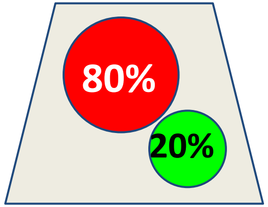
Bayes' Theorem

- Pr(H | E) = Pr(H)Pr(E | H) / [Pr(H)Pr(E | H) + Pr(~H)Pr(E | ~H)]

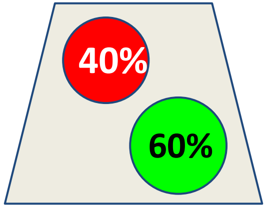
"prior" "true positives" "false positives"

- To use this formulation of Bayes' Theorem in order to tell you how to update your confidence in a hypothesis given a piece of evidence… *you need the rates of the **prior**, **true positives**, and **false positives**.*

Urn A



Urn B



Setup: Pick a ball at random. Then replace and shuffle.

R = Pull a red ball, **A** = Pull from A, **B** = Pull from B

Notice **Pr(R | A)** = .8, **Pr(R | B)** = .4, and **Pr(A)** = **Pr(B)** = .5

Suppose you pull a red ball. What's the probability it came from A?

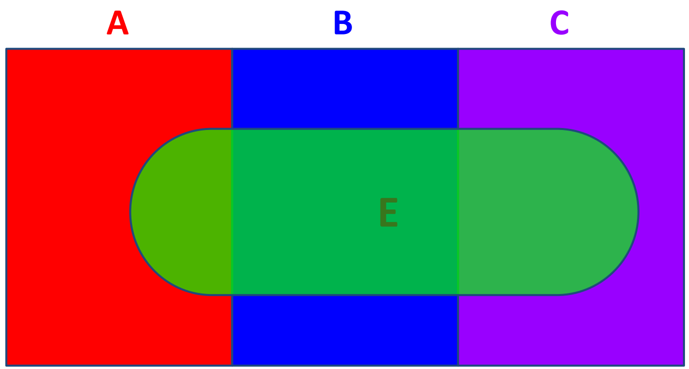
$$\Pr(A \mid R) = \frac{\Pr(A)\Pr(R \mid A)}{\Pr(A)\Pr(R \mid A) + \Pr(\sim A)\Pr(R \mid \sim A)}$$

$$= \frac{(.5 \times .8)}{(.5 \times .8) + (.5 \times .4)}$$

$$= \tfrac{2}{3}$$

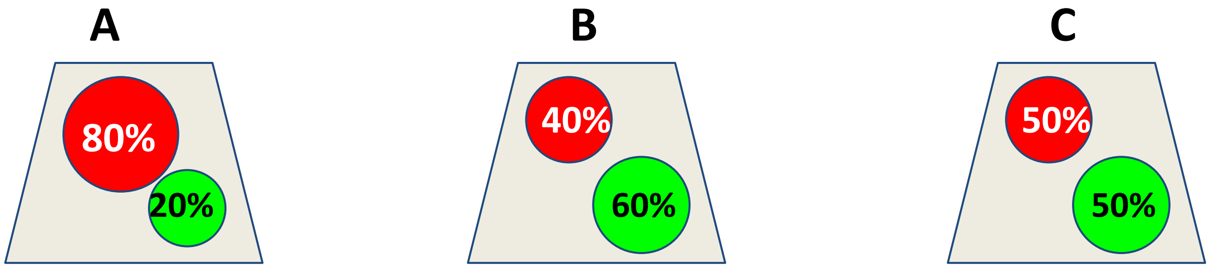
Bayes' Rule

- Note that Bayes' Rule generalizes to cases of multiple hypotheses, if the hypotheses are *mutually exclusive* and *jointly exhaustive*.
  - *mutually exclusive* means only one can be true
  - *jointly exhaustive* means at least one must be true



$$\Pr(A \mid E) = \frac{\Pr(A)\Pr(E \mid A)}{\Pr(E)}$$

$$= \frac{\Pr(A)\Pr(E \mid A)}{\Pr(A)\Pr(E|A) + \Pr(B)\Pr(E|B) + \Pr(C)\Pr(E|C)}$$



Setup: Select an urn randomly. Pick a ball randomly from it. Then replace and shuffle. You pick a red ball. What's the probability it came from A?

$$\Pr(A \mid R) = \frac{\Pr(A)\Pr(R \mid A)}{\Pr(A)\Pr(R \mid A) + \Pr(B)\Pr(R \mid B) + \Pr(C)\Pr(R \mid C)}$$

$$= \frac{(⅓ \times .8)}{(⅓ \times .8) + (⅓ \times .4) + (⅓ \times .5)}$$

$$= \frac{(⅓ \times .8)}{⅓ \times (.8+.4+.5)}$$

$$= .8/1.7$$

$$= 47\%$$

Bayes’ Rule

- 3% of bananas from Honduras have tarantulas on them.
- 6% of bananas from Guatemala have tarantulas on them.
- 40% of our bananas are from Honduras.
- 60% of our bananas are from Guatemala.

A tarantula was found on a randomly selected banana.
What’s the probability it came from Guatemala?
G = Banana came from Guatemala, **H** = Banana came from Honduras, **T** = Banana had a tarantula on it. What is **Pr(G | T)**?

Pr(G | T) = Pr(G)Pr(T | G) / [**Pr(G)Pr(T | G) + Pr(H)Pr(T | H)**]

= (.06 x .6) / [(.6 x .06) + (.4 x .03)]

= 3/4

Odd Question #5

- **85% of taxis are green**. **15% of taxis are blue**.
- A witness testifies that she saw the hit-and-run taxi and that it was blue.
- **This witness is correct 80% of the time** when presented with a series of green and blue cars under the same conditions of the accident.

What's the probability that the hit-and-run taxi was blue *given* the witness's report?

1. 80%
2. Between 50% – 80%
3. 50%
4. Under 50%

Odd Question #5

G = The hit-and-run taxi is green.
B = The hit-and-run taxi is blue.
W_b = The witness says that the hit-and-run taxi is blue.

Note that **Pr(G)** = .85, **Pr(B)** = .15, **Pr(W_b | B)** = .8, **Pr(W_b | G)** = .2

(Note that we *cannot* immediately reason: **Pr(B | W_b)** = .8)

Pr(B | W_b) = **Pr(B)Pr(W_b | B)** / [**Pr(B)(W_b | B)** + **Pr(G)(W_b | G)**]

= (.15 x .8) / [(.15 x .8) + (.85 x .2)]

= 12/29 = **.41**

4. Under 50%. It's more likely that the hit-and-run taxi was green!

Base Rates

- This is an interesting situation. *Our witness* is “reliable” in the sense that she is correct 80% of the time. So, how could *her testimony* be so *un*reliable (worse than a coin toss!)
- Answer: **Our reliable witness is testifying in a situation with a *very skewed base rate – 85% of the taxis are green!***
- Even though she testified that it was blue, it is *more likely* that it was a green taxi that she misidentified vs. a blue taxi that she correctly identified…largely because there are so many more green taxis around!

Base Rates

- The same situation crops up in the context of medical testing.
- Suppose a test for the deadly *Logicavirus* is 99% accurate in this sense:
  - If you have Logicavirus, the test will say YES 99% of the time.
  - If you don't have Logicavirus, it will say NO 99% of the time.
- Suppose 1 in 10,000 people have Logicavirus, and we test 1,000,000 people.
- Then 100 out of those million people have it, and 99% (99 ppl) test positive.
- But 999,900 of those million don't, and 1% (~10,000) test positive.
- So suppose you test positive. What's probability that it's a false positive?

L = You have Logicavirus, **P** = You test positive

Pr(~L | P) = Pr(~L)Pr(P | ~L) / [**Pr(~L)Pr(P | ~L) + Pr(L)Pr(P | L)**]

= (9,999/10,000 x .01) / [(9,999/10,000 x .01) + (1/10,000 x .99)]

= .009999 / [.009999 + .000099] = **.9893...** **(about 99%)**

- This is why "reliable" tests can be misleading *if the disease is rare*. A reliable test can only be trusted when applied to a population "at risk" (cf. the HIV epidemic in 80s)

I give you the rest of our time, if there is any, to working on the problem set and asking myself and Andrew any questions.

Thanks!