this, next time: (see syllabus)

reading: see syllabus

(AMS 131 11 Apr 19)

1

per T. Bayes (1760)

possible causes
- bad water
- bad air
- bad food
- disease(s)
- ⋮

deterministic

causality:

probabilistic causality

village

some people (data) dying

cause → effect ?

unknown U, data D

$p(\text{cause} \mid \text{effect}) = ?$ hard(er)

$p(\text{effect} \mid \text{cause}) = \text{easy}$ (easier)

$p(U \mid D) = ?\, p(D \mid U)$ $\qquad p(D > 0)$

$$p(U \mid D) = \frac{p(U \text{ and } D)}{p(D)} \quad \text{(\&)}$$

(*) = (**)

$\rightarrow p(U \text{ and } D)^{(**)} = p(D) \cdot p(U \mid D)$

$$p(D|u) = \frac{p(D \text{ and } u)}{p(u)} \rightarrow \quad p(D \text{ and } u) = p(u)\,p(D|u)$$ ②

---

therefore $p(u)\,p(D|u) = p(D)\,\underline{p(u|D)}$

$$p(u|D) = \frac{p(u)\,p(D|u)}{p(D)}$$

$$\underbrace{p(\text{unknown} \mid \underline{\text{data}})}_{\text{posterior info}} = \frac{\overbrace{p(\text{unknown})}^{\text{prior info}} \cdot \overbrace{p(\text{data} \mid \text{unknown})}^{\text{data info}}}{\underbrace{p(\text{data})}_{\text{hard}}}$$

<u>Bayes's Theorem</u>

for T/F propositions

$$p(\overset{T/F}{u}|D) = \frac{p(u)\,p(D|u)}{p(D)}$$

$$p(\text{not}\,u|D) = \frac{p(\text{not}\,u)\,p(D|\text{not}\,u)}{p(D)}$$

PLAN AHEA

$$\left[\frac{P(u \mid D)}{P(\text{not } u \mid D)}\right] = \left[\frac{P(u)}{P(\text{not } u)}\right] \cdot \left[\frac{P(D \mid u)}{P(D \mid \text{not } u)}\right]$$

***

(posterior odds ratio given D) = (prior odds ratio in favor of U) (Bayes factor) $\boxed{\dagger}$

$P(H) = p$

$P(T) = 1 - p$

$= P(\text{not } H)$

$$\frac{P(H)}{P(\text{not } H)} = \frac{p}{1-p} = \text{odds ratio in } (O) \text{ favor of } H$$

*** = odds ratio against T

odds ratio against H $= \dfrac{1-p}{p}$

Bayes's Theorem in odds form

time →

a priori — before info

(new) data D arrives — info

a posteriori — after info

$\boxed{\dagger}$ likelihood ratio

$$o = \frac{p}{1-p} \leftrightarrow p = \frac{o}{1+o}$$

Case study: Monte Hall problem

Y_i = { you initially choose door i }

M_j = { Monte Hall then opens door j } (goat)

C_k = { car actually behind door k }

$i, j, k = 1, 2, 3$

without loss of generality

you pick door 1 (Y_1) & Monte opens door 2 to reveal a goat (M_2)

PLAN AHEA

(11)

we want to compare $P(C1 \mid M2, Y1)$ and with $P(C3 \mid M2, Y1)$.

This is like ELISA:

	~~Monte~~	ELISA
unknown:	location of car	true HIV status
data:	Monte showing you a goat behind door 2	what ELISA said

we want $\underline{P(\text{unknown} \mid \text{data})}$ but problem setup gave us $\underline{P(\text{data} \mid \text{unknown})}$

So let's use Bayes's Theorem to reverse order of conditioning in odds form

~~posterior odds~~ prior odds

$$\frac{P(C1 \mid M2, Y1)}{P(C3 \mid M_2, Y_1)} = \left[\frac{P(C1)}{P(C3)} \right] \cdot \left[\frac{P(M2, Y1 \mid C1)}{P(M2, Y1 \mid C3)} \right]$$

Bayes factor

$$P(C2 \mid M2, Y1) = 0$$

Now by the rules $P(c1) = P(c3) = \frac{1}{3}$,

so the prior odds are $\dfrac{P(c1)}{P(c3)} = \dfrac{1/3}{1/3} = 1$

---

to evaluate probabilities like $P(m2\underset{\text{and}}{,}\, y1 \mid c1)$, let's use the general form of product rule for and:

$$\frac{P(m2, y1 \mid c1)}{P(m2, y1 \mid c3)} = \frac{\overset{1}{\cancel{P(y1 \mid c1)}} \cdot P(m2 \mid y1, c1)}{\cancel{P(y1 \mid c3)} \cdot P(m2 \mid y1, c3)}$$

but Y_i and C_j are independent so

$P(y1 \mid c1) = P(y1) = \frac{1}{3}$ and

$P(y1 \mid c3) = P(y1) = \frac{1}{3}$ so

$$\frac{P(c1 \mid m2, y1)}{P(c3 \mid m2, y1)} = \frac{P(m2 \mid y1, c1)}{P(m2 \mid y1, c3)} = \frac{1/2}{1} = \frac{1}{2}$$

So: after (A2) (given Y1, C3),

the posterior odds in favor of

car behind door 3 are 2 : 1,

so $P(C3 \mid A2, Y1) = \frac{2}{3}$ &

you should switch.

Case study: Cromwell's Rule

---

for any D such that $P(D) > 0$

& any A,

Dennis Lindley

(a) if $P(A) = 0$ then $P(A \mid D) = 0$

and

(b) if $P(A) = 1$ then $P(A \mid D) = 1$

D = data	$P(A)$ prior information about A
A = unknown	$P(A \mid D)$ posterior info about A

meaning | Anything you put prior probability 0 $\boxed{1}$ on has to have posterior probability 0 $\boxed{1}$ no matter how the dataset comes out; this destroys the possibility of learning from data.

(a) $$P(A \mid D) = \frac{P(A \text{ and } D)}{P(D)}$$

but if $P(A) = 0$ $(\emptyset)$

then $P(A \cap D) = 0$ ✓

$$P(A \mid D) = \frac{P(A \text{ and } D)}{P(D)}$$

$P(A) = 1$

$A \equiv S$

(9.50)

so $(A \cap D) = D$

and $P(D) / P(D) = 1$

Case Study: The Rasmussen Report 15

WASH-1400, "The Reactor Safety Study"

problem: Estimate P(catastrophic accident at nuclear power plant) (*) per year

at a moment in history (1975) when no such events had ever occurred (nuclear power began in about 1955, ...)

3 mile Island 1979

"Solution": Use expert judgment to break down (*) into a collection of simpler events connected together with (and), (or), ...; for example

$$\longleftarrow \text{lots} \longrightarrow$$

(*) = (thing 1 breaks & alarm 1 fails to go off & thing 2 then breaks & ...)

Result | Their Estimate of $P(\circledast)$ was extremely small : 10^{-12}, & yet only 4 years later : 3 mile Island

---

What went wrong? | Right calculation :

$$P(\circledast) = P\begin{pmatrix}\text{thing 1}\\\text{breaks}\end{pmatrix} \cdot P\left(\begin{matrix}\text{alarm}\\\text{1 fails}\end{matrix} \middle| \begin{matrix}\text{thing}\\\text{1}\\\text{breaks}\end{matrix}\right)$$

$$\cdot P\left(\begin{matrix}\text{thing}\\\text{2 breaks}\end{matrix} \middle| \begin{matrix}\text{thing 1}\\\text{breaks}\end{matrix} \;\&\; \begin{matrix}\text{alarm}\\\text{1 fails, \&}\end{matrix}\right) \cdot \ldots$$

What they did instead: they assumed independence!

$$P(\circledast) = \underbrace{P\begin{pmatrix}\text{thing 1}\\\text{breaks}\end{pmatrix}}_{\text{small}} \cdot \underbrace{P\begin{pmatrix}\text{alarm 1}\\\text{fails}\end{pmatrix}}_{\text{small}} \cdot \underbrace{P\begin{pmatrix}\text{thing 2}\\\text{breaks}\end{pmatrix}}_{\text{small}} \cdots$$

= tiny just because many numbers close to 0 multiplied