this time: conditional probability

next time: discrete distributions

read DS ch. 1, 2, 1st half of ch. 3

office hours now posted on course web page

village | some people dying. why?

cause → effect

A

exhaustive list of possible causes:

- bad water
- bad air
- bad food
- disease

B_i

$P(A \mid B_i) = P(\text{effect} \mid \text{cause})$ easier

or $P(B_i \mid A) = P(\text{cause} \mid \text{effect})$ harder but the goal

Bayes wanted to reverse order of conditioning:

②

cause	effect
B_i	A
unknown	data

$\boxed{\text{thm:}}$ optimal information processing

$$\underbrace{p(\text{cause} \mid \text{effect})}_{(*)} = p(\text{unknown} \mid \text{data}) = \frac{\overset{(**)}{p(\text{unknown})}\ \overset{(***)}{p(\text{data} \mid \text{unknown})}}{p(\text{data})}$$

(a priori) before data — after (a posteriori)

time →

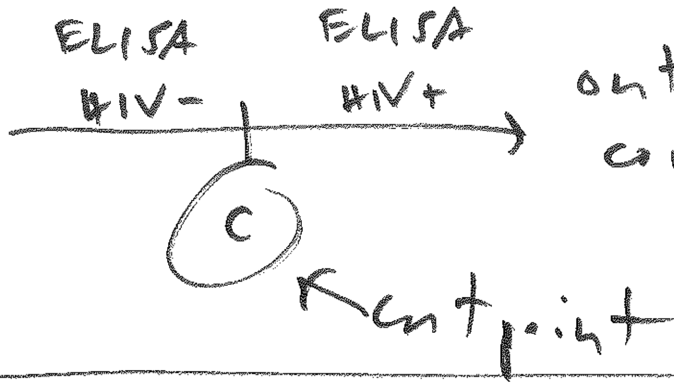
info. about unknown | data arrives | info. about unknown

prior info. | data info | posterior info.

$$\overset{(*)}{\left(\begin{matrix}\text{posterior}\\ \text{info}\end{matrix}\right)} = \frac{\overset{(**)}{\left(\begin{matrix}\text{prior}\\ \text{info}\end{matrix}\right)} \cdot \overset{***}{\left(\begin{matrix}\text{data info}\\ \text{[likelihood]}\end{matrix}\right)}}{\text{(annoying) normalizing constant}}$$

LTP $P(A) = \sum_{i=1}^{k} P(B_i)\, P(A \mid B_i)$

$\therefore \quad P(B_i \mid A) = \dfrac{P(B_i) \cdot P(A \mid B_i)}{\sum_{j=1}^{k} P(B_j) \cdot P(A \mid B_j)}$



$A = \{\text{true HIV+}\}$

$\text{not } A = \{\text{true HIV−}\}$

$\oplus = \{\text{ELISA says } +\}$

$\ominus = \{\text{———— } -\}$

sensitivity = 0.96

$P(+ \mid A)$

specificity = 0.97

$P(- \mid \text{not } A)$

prevalence =

$P(A) = 0.4\%$

$= 0.004$

[Suppose ELISA says (+) for a given blood sample: what is $P(A \mid +)$? 4]

method 1: Bayes's Thm directly

$$P(A \mid +) = \frac{\underset{\substack{.004 \\ \text{prevalence}}}{P(A)} \cdot \underset{\substack{\text{sensitivity} \\ .96}}{P(+ \mid A)}}{\underset{\text{annoying normalizing constant}}{P(+)}}$$

we want $P(+)$

this is hard to think about unless we know $\boxed{\text{truth}(A, \text{not}A)}$,

because we do know $P(+ \mid A)$

and we can work out $P(- \mid A)$

from that & also $P(+ \mid \text{not}A)$

$= 1 - P(- \mid \text{not}A) = 1 - \text{specificity} = .03$

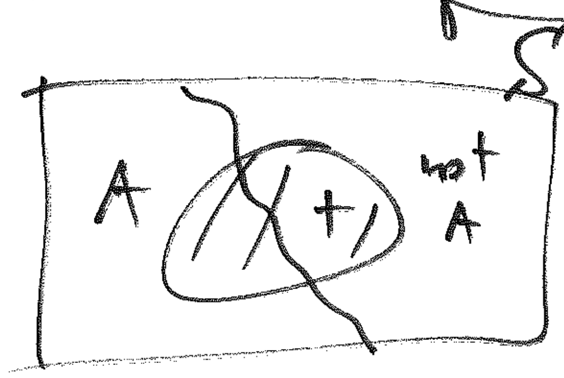
when looking for a partition, Dennis Lindley (1920 – 2014) suggests:

$$\begin{pmatrix}\text{finding} \\ \text{a} \\ \text{partition}\end{pmatrix} = \begin{pmatrix}\text{"extending} \\ \text{the conversation"}\end{pmatrix}$$

a useful partition $B_1, \dots, B_k$ has 2 properties: $P(B_i) > 0$ & known & $P(+ \mid B_i)$ is also known

---

I don't know $P(+)$ but I do know something about $P(\underset{\text{data}}{+} \mid \underset{\text{unknown truth}}{A})$ and $P(+ \mid \text{not } A)$



$$P(+) = P(A)\cdot P(+|A) +$$

$$P(\text{not } A)\cdot P(+|\text{not } A)$$

$$= (.004)(0.96) + (.996)(.03)$$

$$\therefore\ P(A|+) = \frac{(.004)(0.96)}{(.004)(0.96) + (.996)(.03)}$$

positive

$$= \frac{384}{3372} \doteq 0.11\ (!)$$

---

false positive rate =

$$1 - P(A|+) = 0.89 \quad \text{(huge)}$$

$$= P(\text{not } A|+)$$

$$P(A|-) = \text{false negative rate}$$

$$= \frac{16}{96628} \doteq .0002 = 0.02\%$$

method 2: 2x2 contingency table

truth

(.961) · 400

(.97) · 99600

what ELISA says	has HIV (A)	no HIV (not A)	
+	384	2,988	3,372
−	16	96,612	96,628
	400	99,600	100,000

sensitivity

specificity

$P(A)$ = prevalence = .004

$$P(A \mid +) = \frac{384}{3372} = 0.11 \; (!)$$

(utility) blood bank

utility function

truth

ELISA	A	not A
+	✓	wasted good unit of blood
−	really bad	✓

(bad) ~ \$50-100

somebody gets HIV from "bad" blood

optimal decision-making

Bayesian decision theory

method 3: odds

if $P(C) = p$ then odds in favor of C are/is $O = \frac{p}{1-p}$

Bayes's thm in odds form

$$\underbrace{\left[\frac{P(A \mid +)}{P(\text{not } A \mid +)}\right]}_{\substack{\text{posterior odds} \\ \text{in favor of } A}} = \underbrace{\left(\frac{P(A)}{P(\text{not } A)}\right)}_{\substack{\text{prior odds in} \\ \text{favor of } A}} \cdot \underbrace{\left(\frac{P(+ \mid A)}{P(+ \mid \text{not } A)}\right)}_{\substack{\text{likelihood ratio} \\ \text{or Bayes factor}}}$$

$+$ ← data

$$P(A \mid +) = \frac{P(A) \cdot P(+ \mid A)}{P(+)}$$

$$P(\text{not } A \mid +) = \frac{P(\text{not } A) \cdot P(+ \mid \text{not } A)}{P(+)}$$

$\frac{96}{3}$

$$\text{posterior odds in favor of } A = \left(\frac{.004}{.996}\right)\left(\frac{0.96}{\underbrace{1-0.97}_{1-\text{spec.}}}\right) \quad \leftarrow \text{sens.}$$

$$\left(\text{prior odds}\right) \qquad \left(\text{Bayes factor}\right)$$

$$\frac{32}{249} = \left(\begin{array}{c}249 \\ \text{to } 1 \\ \text{against} \\ \text{really HIV+}\end{array}\right) \cdot \left(\begin{array}{c}32 \text{ to } 1 \\ \text{in favor} \\ \text{of really HIV+}\end{array}\right)$$

$$o = \frac{p}{1-p} \leftrightarrow p = \frac{o}{1+o} = \frac{\frac{32}{249}}{1+\frac{32}{249}} \doteq 0.11$$

$$= \frac{384}{3372} = \frac{32}{281}$$

Monte Hall problem

ELISA : ~~$~~ ~ $10

fp rate 89% (0.5%) fn rate .02% (.001%)

western blot ~ $100

Case study:

Monte Hall problem

Y_i = { you initially choose door i }

M_j = { Monte Hall then opens door j }

C_k = { car actually behind door k }

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

without loss of generality you pick door 1 (Y_1) & Monte (M_2) opens door 2 to reveal a goat

PLAN AHEA

we want to compare $P(C1 \mid M2, Y1)$ and with $P(C3 \mid M2, Y1)$. 11

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

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

$$\frac{P(C1)}{P(C3)} = \frac{1/3}{1/3} = 1$$

---

to evaluate probabilities like $P(M2, Y1 \mid C1)$, let's use the

and

general form of product rule for and:

$$\frac{P(M2, Y1 \mid C1)}{P(M2, Y1 \mid C3)} = \frac{\cancel{P(Y1 \mid C1)} \cdot P(M2 \mid Y1, C1)}{\cancel{P(Y1 \mid C3)} \cdot P(M2 \mid Y1, C3)}$$

1

but Y_w 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 h_2 (given y_i, c_j), the posterior odds in favor of car behind door 3 are 2 : 1,

so $P(C_3 \mid h_2, y_1) = \frac{2}{3}$ & you should switch.

Case study: Cromwell's Rule

Dennis Lindley

for any D such that $P(D) > 0$ & any A,

(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 $\boxed{1}$ 0 on has to have posterior probability $\boxed{1}$ 0 no matter how ~~bad~~ the dataset comes out; this destroys the possibility of learning from data.

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

but if $P(A) = 0$ (ϕ)

then $P(A \cap D) = 0$ ✓

$$P(A \mid D) = \frac{P(A \text{ and } D)}{P(D)} \qquad P(A) = 1$$

A is S

so $(A \cap D) = D$

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

(9.50)

Case Study: The Rasmussen Report

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

⊛ = (thing 1 breaks & alarm 1 fails to go off & thing 2 breaks & ...) ← lots →

Result | The Estimate of $P(\circledast)$ was extremely small: 10^{-12}, &

yet only 4 years later: 3 mile Island

What went wrong? | Right calculation:

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

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

What they did instead: they assumed independence!

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

= tiny just because many numbers close to 0 multiplied