Case Study (CS) 1 (HIV screening) (Uncertainty about the truth of a true-false proposition) ①

You're a doctor; a new (to you) patient, Bob, comes to you wondering if he is HIV positive.

~~Notation~~/ Definition: θ = something both interesting & unknown to you (You [Good, 1950]: a person wishing to sensibly reason in the presence of uncertainty). Def: Uncertainty: a state of incomplete information about something of interest to You (e.g., θ)

Notation: Θ (capital θ) = the set of possible values of θ

CS 1: $\theta = \begin{cases} 1 & \text{if Bob is HIV+} \\ 0 & \text{— is not —} \end{cases}$

so here $\Theta = \{0, 1\}$.

often Θ = the real number line $\mathbb{R}$

or a subset of $\mathbb{R}$; later we'll work on problems in which Ⓗ $= \mathbb{R}^k$ for k (integer) > 1; in principle Ⓗ could be almost anything; later we'll examine problems with Ⓗ $=$ {all continuous CDFs F on $\mathbb{R}$}

cumulative distribution function

this is an uncountably infinite set

(CS1, continued) You take a history from Bob, gathering (background) information that You judge to be relevant to Bob's HIV status; e.g.: Bob is male, 28 years old, is gay, mostly (but not always) practices safe sex ⊛

Def. $\mathcal{B} = \{ B_1, \dots, B_b \}$ (b a positive finite integer) is a set of (true/false) propositions, all rendered true by the background context of the problem ⊛⊛

CS1: $B_1 =$ (Bob is male), $B_2 =$ (Bob is 28 years old), ..., $B_4 =$ (Bob ⊛).

(**) and intended to be exhaustive of all relevant background contextual information.

Axiom 1 Your uncertainty about the truth of a proposition A given background information $\mathcal{B}$ can be quantified with the conditional probability $P(A \mid \mathcal{B})$.

CS1 Here this is $p_0 = P(\theta = 1 \mid \mathcal{B}) = P(\text{Bob is HIV+} \mid \mathcal{B})$.

This can be quantified through activities such as elicitation of expert judgment, web searches of relevant publications, ... The CDC estimates the prevalence of HIV in Americans of age 18+ at 0.3%, but Bob's background information includes risk factors implying that a higher value of p_0 is appropriate for him. To have a realistic number to work with, let's take $p_0 = 0.01 = 1\%$. This quantifies Your/Bob's uncertainty when he walked into your office, but Bob is hoping You can do better than that, by gathering <u>data</u> uniquely specific to him via a blood test for HIV. Many such tests are now available; we'll look at one called "Determine HIV 1/2 rapid", which returns results in about 10 minutes at a cost of only about US $0.20. Suppose that

www.ncbi.nlm.nih.gov/pmc/articles/PMC88030

this test comes back positive for HIV on Bob's blood; what is Your new estimate of the probability that he really is HIV+?

Axiom 2 / Def. / Notation You will typically have sufficient time/money resources to gather a <u>data set</u> D to update Your uncertainty in light of new <u>relevant</u> <u>information</u> in D.

In simple problems D will consist of a data vector of real numbers $y = (y_1, \dots, y_n)$, $y_i \in \mathbb{R}$ (n a positive finite integer); in more complicated problems D could be almost anything, but will often be a matrix of real numbers.

Let $\mathbb{D}$ = {all possible datasets in the current problem}

CS 1 Let $y_1 = \begin{cases} 1 & \text{if blood test } (+) \\ 0 & \text{—} \; (-) \end{cases}$

then in CS 1, $D = (y_1)$, and our updated probability must be $P(\theta = 1 \mid y_1 = 1, \text{ and } \mathcal{B})$

(Here $\mathbb{D} = \{0, 1\}$) (see Axiom 3 below) (p. 7.1)

Big Picture What we've done so far is a special case of the following paradigm:

---

1. You start with a <u>problem</u> $\mathbb{P}$ involving uncertainty.
2. You notice that $\mathbb{P}$ can be decomposed as follows: $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$,

in which $\mathbb{Q}$ is the (finite) set of real-world questions to be answered & $\mathbb{C}$ is the real-world <u>context</u> in which these questions arise.

(3) $\mathbb{Q}$ and $\mathbb{C}$ jointly specify θ, the unknown of main interest, and D, the dataset (<u>information</u> resource) that You can use to update your uncertainty given (in light of) all relevant information available to You. (4) $\mathbb{C}$ specifies $\mathcal{B} = \{ B_1, \dots, B_b \}$, which captures all relevant background (contextual) information in $\mathbb{P}$. Symbolically, what has happened so far is

lives in Θ lives in $\mathbb{D}$

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \longrightarrow (\theta, D, \mathcal{B}) \qquad \mathcal{B} \to \{ B_1, \dots, B_b \}$$

natural language | math

I've created this formalism to (attempt to) capture what goes on in the problem-formulating step of problem-solving involving uncertainty. (7)

Draper (2019) | $\boxed{\text{Advice}}$ If we wish to use (applied) math to solve real-world problems, we must move from the world of natural language to the world of math as fast as possible without omitting any relevant information; this formalism does that.

CS 1 | $\boxed{\mathbb{P} = (\mathbb{Q}, \mathbb{C})}$ $\mathbb{Q}$ = (what is the current best-information estimate of the probability that Bob is HIV positive?)

$\mathbb{C} \to \mathcal{B} = \{ B_1, \dots, B_4 \}$ (stated previously on p. (2))

$$\theta = \begin{cases} 1 & \text{if Bob really is HIV+} \\ 0 & \text{------------------ HIV-} \end{cases}$$

$\Theta = \{0, 1\}$

$$D = (y_1), \quad y_1 = \begin{cases} 1 & \text{if blood test says HIV+} \\ 0 & \text{------------------ HIV-} \end{cases}$$

$\mathbb{D} = \{0, 1\}$

Axiom (3) here (p. (2.1))

we want to update from $P(\theta = 1 \mid \mathcal{B})$ to $P(\theta = 1 \mid y_1 = 1, \mathcal{B})$.

Axiom 3 Suppose, following Axiom 1, that Your current information (uncertainty) about the truth of a proposition A, given that the totality of Your current (relevant) information (uncertainty) about A is summarized by the (true) proposition B, is quantified with $P(A|B)$. If new information in the form of a proposition C (judged by You to be both accurate (true) and relevant) arrives, Your updated information (uncertainty) about A is optimally quantified with $P(A|BC)$.

C can (of course) be a compound proposition of the form $C = (C_1 \text{ and } C_2 \text{ and} \ldots \text{and } C_k) \triangleq (C_1, C_2, \ldots, C_k) \triangleq (C_1 C_2 \cdots C_k)$

This is a special case of updating from $P(A|C)$ to $P(A|B,C)$, in which A, B, C are all propositions. If you don't know how to do something mathematical, just start playing around with it; maybe something good will pop out.

Advice

Reminder Def.

$$P(A|C) = \begin{cases} \dfrac{P(AC)}{P(C)} & \text{if } P(C) > 0 \\ \text{undefined} & P(C) = 0 \text{ (i.e., } C \text{ is false)} \end{cases}$$

same as $P(A,C)$

" $P(A \text{ and } C)$

deMoivre (1710)
Bayes (1763)

Advice Try hard not to condition on false propositions; that's the probability equivalent of dividing by 0 in the real-number system. This can sometimes be done in a meaningful manner, but it requires careful analytic manipulation.

Let's play

You will quickly find that

$$P(A|BC) = P(A|C) \pm \boxed{?}$$

leads nowhere useful, so let's try

Assume all denominators are positive

$$P(A|BC) = P(A|C) \cdot \boxed{?}$$

$$\frac{P(ABC)}{P(BC)}^{\boxed{1}} = \frac{P(AC)}{P(C)}^{\boxed{2}} \cdot \boxed{\frac{P(ABC)^{\boxed{3}} \cdot P(C)^{\boxed{4}}}{P(AC) \cdot P(BC)}}$$

This hits paydirt: rewriting in terms of conditional probabilities, we have just proven

Theorem (a version of Bayes's Theorem for propositions): if A, B, and C are all non-false propositions,

$$P(A|BC)^{\boxed{1}} = P(A|C)^{\boxed{2}} \cdot \frac{P(B|AC)^{\boxed{3}}}{P(B|C)_{\boxed{4}}}$$

10

$$\underbrace{P(A \mid BC)}_{\boxed{1}} = \frac{\overset{\boxed{2}}{P(A \mid C)} \cdot \overset{\boxed{3}}{P(B \mid AC)}}{\underbrace{P(B \mid C)}_{\boxed{4}}}$$

in a slight abuse of notation:

this is Bayesian updating of information

$$\underbrace{P(\theta \mid D B)}_{\boxed{1}\ ***} = \frac{\underbrace{P(\theta \mid B)}_{\boxed{2}}\ \underbrace{P(D \mid \theta B)}_{\boxed{3}}}{\underbrace{P(D \mid B)}_{\boxed{4}}}$$

("posterior" information about the unknown θ given data D & context B) = $\dfrac{(\text{"prior" information about } \theta \text{ given context } B) \cdot (\text{likelihood}^{*} \text{ information about } \theta \text{ given } D \text{ and } B)}{(\text{normalizing constant})^{**}}$

(data)

---

* This is ~~not~~ yet obvious but will become so in Case Study 2

** Since *** is a function of θ for fixed D and B, $P(D \mid B)$ is constant in θ

(10.1)

Terminology | The presence of the dataset D creates an information dichotomy (partition):

Your total information about θ $\boxed{1}$

information about θ internal to D | information about θ external to D $\boxed{2}$

As we'll see more clearly in Case Study 2, this is the likelihood information about θ given D and $\mathcal{B}$ $\boxed{3}$.

This should be called the external information about θ given $\mathcal{B}$, but a different name has been embedded so deeply in the Bayesian story (1960) that it's too late to change it:

often (but not always) Your external information arrives before the dataset:

time →

information about θ before D arrives | D arrives | information about θ after D arrives

latin: a priori (prior) $\boxed{2}$ | latin: a posteriori (posterior) $\boxed{1}$

CS 1 $$p(\theta=1 \mid y_1=1, \mathcal{B}) = \frac{p(\theta=1 \mid \mathcal{B})\, p(y_1=1 \mid \theta=1, \mathcal{B})}{p(y_1=1 \mid \mathcal{B})} \quad (1)$$

we have already established that the prior information estimates $p(\theta=1 \mid \mathcal{B}) = p_0$ for Bob is $0.01 = 1\%$; what about the other two probabilities?

$$\boxed{p(y_1=1 \mid \theta=1, \mathcal{B})}$$

in words, this is $p\left(\begin{array}{c}\text{blood test} \\ \text{says HIV+}\end{array} \,\middle|\, \begin{array}{c}\text{really is} \\ \text{HIV+}, \cancel{\mathcal{B}}\end{array}\right)$ (****)

$\mathcal{B}$ only describes background information about Bob & is therefore ~~irrelevant~~ in

(or could put information about blood test in $\mathcal{B}$: better)

(****) is a standard measure of the accuracy of a blood test called the sensitivity of the test; for the "Determine HIV" test this probability is estimated to be 0.999

(excellent)

$\boxed{P(Y_1 = 1 \mid \mathbb{B})}$ Bayesians sometimes refer

to this as the <u>annoying</u> normalizing constant, because it's often the hardest ingredient in Bayes's Theorem to pin down. I will

now show you 3 different ways to estimate this probability, all of them

instructive:
(8 Jan 19)

Method 1: 2x2 <u>contingency</u> table

Imagine randomly sampling 100,000 people blood from the <u>population</u> $\mathcal{P}$ = {all people similar to Bob in all relevant ways} (i.e., people with the same $\mathbb{B}$) and running all the blood samples through "Determine HIV"; cross-tabulate (truth) against (what blood test says), by which I mean what we <u>expect</u> to happen: