

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

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

$$CS 1: \theta = \left\{ \begin{array}{l} 1 \text{ if Bob (is) HIV+} \\ 0 \text{ — (is not) —} \end{array} \right\} \text{ often}$$

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

\mathbb{H} = the real number line \mathbb{R}

or a subset of \mathbb{R} ; later we'll work on problems ⁽²⁾

in which $\mathcal{H} = \mathbb{R}^k$ for k (integer) > 1 ; in

principle \mathcal{H} could be almost anything; later we'll examine ^{problems with} $\mathcal{H} = \{ \text{all continuous CDFs } F \text{ on } \mathbb{R} \}$

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. $B = \{ B_1, \dots, B_b \}$

(b a positive finite integer) is a set of (true/false) propositions, all rendered true by the background context of the problem ^{**}

CS1: $B_1 = (\text{Bob is male})$

$B_2 = (\text{Bob is 28 years old}), \dots, B_4 = (\text{Bob } \oplus)$.

~~1~~ and intended to be exhaustive of all ③
relevant background contextual information.

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

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

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 at 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 data 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~~ 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 data set D to update your uncertainty

in light of new relevant information in D. (5)

In simple problems D will consist of a data
of real numbers
vector $y = (y_1, \dots, y_n)$, (n a positive finite integer),
 $y_i \in \mathbb{R}$

in more complicated problems D could be
almost anything, but will often be a matrix
of real numbers.

let $D = \{\text{all possible datasets in the current problem}\}$
CS 1 let $y_1 = \begin{cases} 1 & \text{if test } (+) \\ 0 & \text{if test } (-) \end{cases}$

then in CS 1, $D = (y_1)$ and our updated
probability must be $P(\theta = 1 \mid y_1 = 1, B)$
(Here $D = \{0, 1\}$) (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 problem P involving uncertainty.
- ② You notice that P can be decomposed as follows: $P = (Q, C)$,

in which \mathcal{Q} is the ^{finite} set of real-world questions to be answered & \mathcal{C} is the real-world context in which those questions arise. ⑥

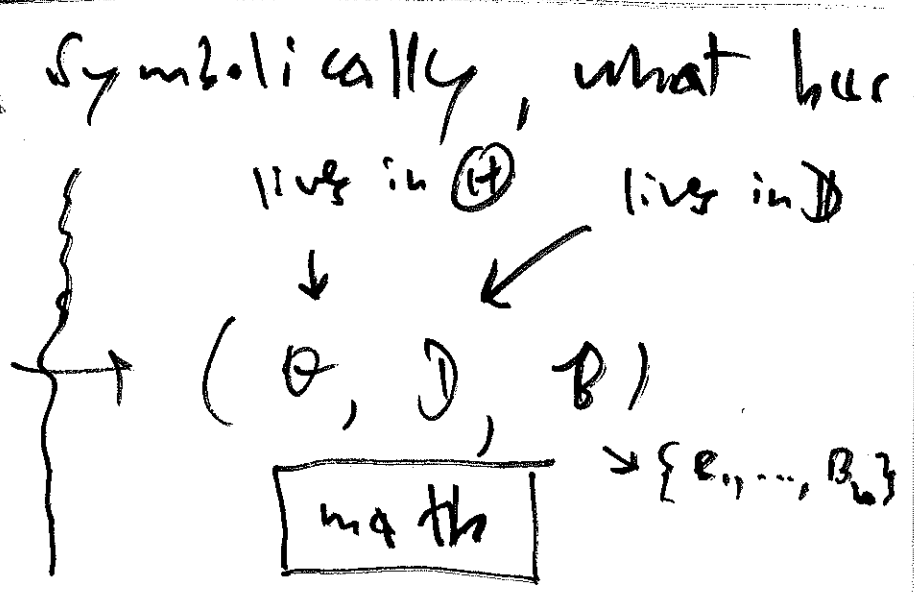
③ \mathcal{Q} and \mathcal{C} jointly specify θ , the unknown of main interest, and \mathcal{D} , the dataset (information resource) that you can use to update your uncertainty given (in light of) all relevant information available to you. ④ \mathcal{C} specifies $\mathcal{B} = \{B_1, \dots, B_b\}$

which captures all relevant background (contextual) information in \mathcal{P} .

happened so far is

$$\mathcal{P} = (\mathcal{Q}, \mathcal{C})$$

natural language



I've created this formalism to (attempt to) $\textcircled{7}$ capture what goes on in the problem-formulation step of problem-solving involving uncertainty.

Draper (2019)

Advice

use (applied) math to solve real-world problems, we must ~~move~~ ^{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 I | $P = (Q, C)$ / $Q =$ (what is the current best-information estimate of the probability that Bob is HIV positive?)

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

$C + B = \{B_1, \dots, B_k\}$
(stated previously) on p. $\textcircled{2}$

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

Axiom $\textcircled{3}$ here (p. $\textcircled{2.1}$)

we want to update from $P(\theta=1|B)$ to $P(\theta=1|y_1=B)$.

Axiom ③ (Suppose, following Axiom ①, that your ^(7.1) 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~~ with $P(A|BC)$.

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

This is a special case of updating from $P(A|c)$ to $P(A|B, c)$, in which A, B, c are all propositions. ⑧

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.

$$P(A|c) = \begin{cases} \frac{P(Ac)}{P(c)} & \text{if } P(c) > 0 \\ \text{undefined} & \text{if } P(c) = 0 \end{cases}$$

de Moivre (1710) Bayes (1763) (i.e., c is false) same as $P(A, c)$
" $P(A \text{ and } c)$ "

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 try

You will quickly find that

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

leads nowhere useful, so let's try

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

Assume all denominators are positive

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

1
2
3
4

This hits pay dirt: 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,

$$P(A|BC) = P(A|C) \cdot \frac{P(B|AC)}{P(B|C)}$$

1
2
3
4

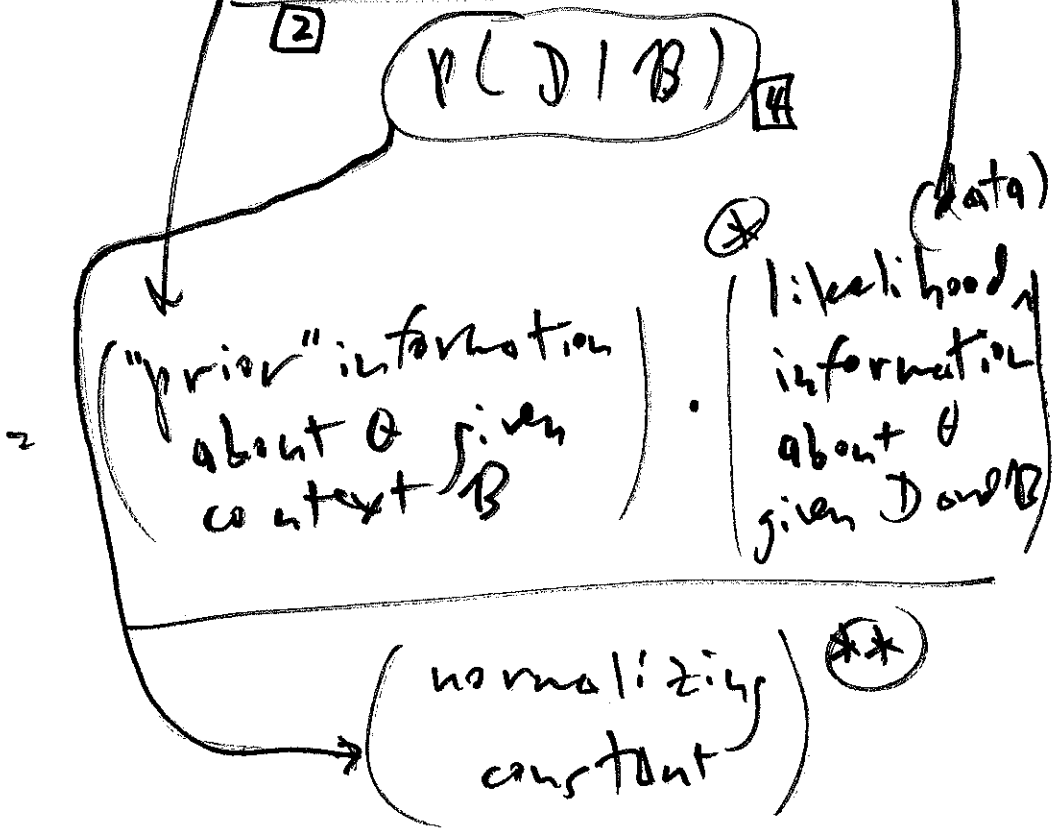
$$P(A|BC) = \frac{P(A|C) \cdot P(B|AC)}{P(B|C)}$$

in a slight abuse of notation:

This is Bayesian updating of information

$$P(\theta | D, B) = P(\theta | B) P(D | \theta, B)$$

"posterior" information about the unknown θ given data D & context B

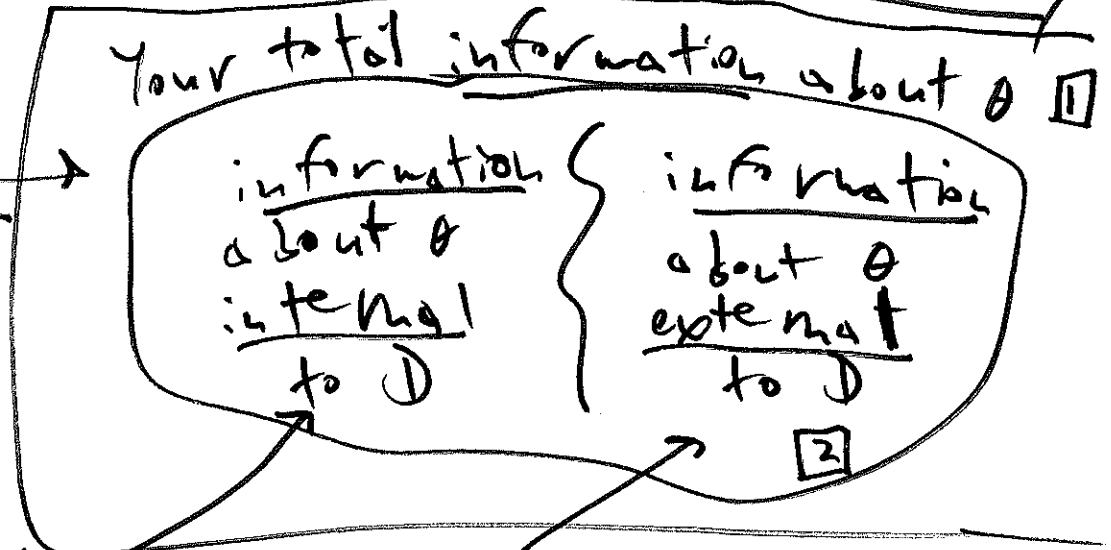


* This is not yet obvious but will become so

in Case Study 2 ** Since ** is a function of θ for fixed D and B , $P(D|B)$ is constant in θ

Terminology | The presence of the dataset

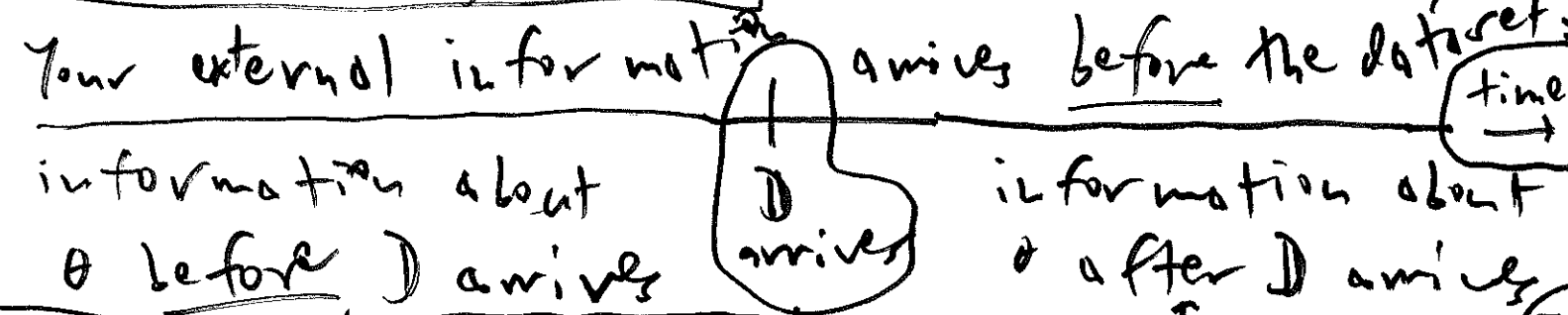
D creates an information dichotomy (partition):



As we'll see more clearly in Case Study 2,

this is the likelihood information about θ given D and B [3].

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



Latin: a priori (prior) [2] | Latin: a posteriori (posterior) [1]

CS1 | $P(\theta=1 | \gamma_1=1, B) = P(\theta=1 | B) P(\gamma_1=1 | \theta=1, B)$

We have already established that the prior

$$P(\gamma_1=1 | B)$$

information estimates $P(\theta=1 | B) = p_0$ for Bob is $0.01 = 1\%$; what about the

other two probabilities?

$$P(\gamma_1=1 | \theta=1, B)$$

in words, this is $P(\text{blood test says HIV+} | \text{really is HIV+})$

B only describes background information about Bob & is therefore ~~irrelevant~~ in (or could put information about blood test in 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~~ ^{be} 0.999 (excellent) ↑

$P(Y_1 = 1 | B)$ Bayesians sometimes refer ⁽¹²⁾ to this as the annoying 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: Method 1: 2x2 contingency table
(8 Jan 19)

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