Ex. 4.1: Show that there is no such nontrivial extension of the binary case. More specifically, prove that if (4.28) and (4.29) hold with $n > 2$, then at most one of the factors

$$\frac{P\left(D_1|H_i\,X\right)}{P\left(D_1\middle|\overline{H_i}\,X\right)}\cdots\frac{P\left(D_m|H_i\,X\right)}{P\left(D_m\middle|\overline{H_i}\,X\right)}$$

is different from unity, therefore at most one of the data sets D_j can produce any updating of the probability for H_i.

Answer:

The equations mentioned are that of independence in D_1 to D_m conditioned to both the hypothesis and its negation:

$$P\left(D_1,\ldots,D_m|A\,X\right)=\prod_j P\left(D_j|A\,X\right)\quad\forall A\in\{H_i,\,\overline{H_i}\}\forall i\in[1,n]_{\mathbb{N}},$$

under the hypothesis that the H_i are exhaustive.

This means

$$\begin{aligned}P(H_i|D_j\,X)+P\left(\overline{H_i}\middle|D_j\,X\right)&=1,\\ \left(P(D_j|H_i\,X)-P\left(D_j\middle|\overline{H_i}\,X\right)\right)P(H_i|X)&=P(D_j|X)-P\left(D_j\middle|\overline{H_i}\,X\right).\end{aligned}$$

I believe the proof has to do with the fact that, for a proposition D_j such that $P(D_j|H_i\,X)\neq P\left(D_j\middle|\overline{H_i}\,X\right)$,

$$P(H_i|X)=\frac{P(D_j|X)-P\left(D_j\middle|\overline{H_i}\,X\right)}{P(D_j|H_i\,X)-P\left(D_j\middle|\overline{H_i}\,X\right)},$$

but so far it's leading me nowhere.

Maybe use that

$$\begin{aligned}P(D_k|\overline{H_i})-P(D_k|H_i)&=\sum_{j\neq i}\frac{P(D_k|H_j)P(H_j)}{P(\overline{H_i})}-P(D_k|H_i)\\&=\frac{1}{P(\overline{H_i})}\sum_j\left[P(D_k|H_j)-P(D_k|H_i)\right]P(H_j),\\ \left[P(D_k|\overline{H_i})-P(D_k|H_i)\right]_k&=\frac{1}{P(\overline{H_i})}\left[P(D_k|H_j)-P(D_k|H_i)\right]_{kj}\left[P(H_j)\right]_j\end{aligned}$$

Ex. 4.2: Calculate the exact threshold of skepticism $f_t(x,\,y)$, supposing that proposition C has instead of 10^{-6} an arbitrary prior probability $P(C|X)=x$, and specifies instead of 99/100 an arbitrary fraction y of bad widgets. Then discuss how the dependence on x and y corresponds - or fails to correspond - to human common sense.

Answer:

Okay, let's do this.

Evidence is defined by

$$\begin{aligned}e(H|DX)&=10\log_{10}O(H|DX)\\&=10\log_{10}\left[\frac{P(H|DX)}{P(\overline{H}|DX)}\right]\\&=10\log_{10}\left[\frac{P(D|HX)P(H|X)}{P(D|\overline{H}X)P(\overline{H}|X)}\right]\\&=e(H|X)+10\log_{10}\left[\frac{P(D|HX)}{P(D|\overline{H}X)}\right].\end{aligned}$$

So, for multiple mutually exclusive and exhaustive hypotheses,

$$e(H_i|DX) = e(H_i|X) + 10\log_{10}\left[\frac{P(D|H_iX)}{P(D|\sum_{j\neq i} H_jX)}\right],$$

and we can replace

$$P\left(D\middle|\sum_{j\neq i} H_jX\right) = P\left(D\sum_{k\neq i} H_j\middle|\sum_{j\neq i} H_jX\right) \tag{1}$$

$$= \sum_{k\neq i} \frac{P(DH_k\sum_{j\neq i} H_j|X)}{P(\sum_{j\neq i} H_j|X)} \tag{2}$$

$$= \frac{\sum_{k\neq i} P(DH_k|X)}{\sum_{j\neq i} P(H_j|X)} \tag{3}$$

$$= \frac{\sum_{k\neq i} P(D|H_kX)P(H_k|X)}{\sum_{k\neq i} P(H_k|X)}. \tag{4}$$

So, letting $w_k^{(i)} = P(H_k|X)/\sum_{j\neq i} P(H_j|X)$,

$$e(H_i|DX) = e(H_i|X) + 10\log_{10}\left[\frac{P(D|H_iX)}{\sum_{k\neq i} w_k^{(i)} P(D|H_kX)}\right].$$

This is, we're comparing the likelihood of H_i with the sum of the likelihoods H_k for $k \neq i$ weighted by their prior likelihood. So yeah, I guess we can keep the jth term of the denominator as long as it's 10 times bigger than the rest of them.

In the problem at hand,

$$\begin{aligned} P(A|X) &= \frac{1}{11}(1-x), \\ P(B|X) &= \frac{10}{11}(1-x), \\ P(C|X) &= x. \end{aligned}$$

Furthermore, if after m measurements we find fm of them are bad,

$$\begin{aligned} P(D|AX) &= \left(\frac{1}{3}\right)^{fm}\left(\frac{2}{3}\right)^{(1-f)m}, \\ P(D|BX) &= \left(\frac{1}{6}\right)^{fm}\left(\frac{5}{6}\right)^{(1-f)m}, \\ P(D|CX) &= y^{fm}(1-y)^{(1-f)m}. \end{aligned}$$

This all results in

$$\begin{aligned} e(C|DX) &= e(C|X) + 10\log_{10}\left[\frac{P(D|CX)}{w_A^{(C)}P(D|AX) + w_B^{(C)}P(D|AX)}\right] \\ &= 10\log_{10}\left[\frac{x}{1-x}\right] + 10\log_{10}\left[\frac{y^{fm}(1-y)^{(1-f)m}}{\frac{1}{11}\left(\frac{1}{3}\right)^{fm}\left(\frac{2}{3}\right)^{(1-f)m} + \frac{10}{11}\left(\frac{1}{6}\right)^{fm}\left(\frac{5}{6}\right)^{(1-f)m}}\right] \\ &= 10\log_{10}\left[\frac{x}{1-x}\right] + 10\log_{10}\left[\frac{\left(\left(\frac{y}{1-y}\right)^f(1-y)\right)^m}{\frac{1}{11}\left(\left(\frac{1}{2}\right)^f\frac{2}{3}\right)^m + \frac{10}{11}\left(\left(\frac{1}{5}\right)^f\frac{5}{6}\right)^m}\right]. \end{aligned}$$

I find the last form more enlightening for the problem at hand. All of those terms are of the form something^m. For $m \to \infty$, the term with the higher base will govern the denominator. To see which, we note the quotient of them is

$$\left(\frac{5}{2}\right)^f \frac{4}{5},$$

so setting this to unity yields a critical value $f = \log\left(\frac{5}{4}\right) / \log\left(\frac{5}{2}\right) \approx 0.24$. Given that the term in parentheses is greater than one, this means hypothesis A governs when f goes above this value, and B governs otherwise. If we want the threshold for $y > \frac{1}{3}$, it is clear that we need to work in the range where hypothesis A is more likely. So the threshold of skepticism is given by the value f such that the quotient

$$\left(2\frac{y}{1-y}\right)^f \frac{3}{2}(1-y)$$

is equal to 1. This yields the exact value

$$f_t(x,\, y) = \frac{\log\left(\frac{2}{3}\frac{1}{1-y}\right)}{\log\left(2\frac{y}{1-y}\right)}.$$

For $y = 99/100$, this yields approximately 0.794155, in conflict with the value of 0.793951 given in the book. But the solution book I'm contrasting with in github gives this same value, 0.7941etc.

The general case for estimating this threshold is now obvious:

$$f_t = \frac{\log\left[\frac{P(\text{good}|CX)}{P(\text{good}|AX)}\right]}{\log\left[\frac{P(\text{bad}|CX)P(\text{good}|AX)}{P(\text{good}|CX)P(\text{bad}|AX)}\right]}.$$

Ex. 4.3: Show how to make the robot skeptical about both unexpectedly high and unexpectedly low numbers of bad widgets in the observed sample. Give the full equations. Note particularly the following: if A is true, then we would expect, according to the binomial distribution (3.86), that the observed fraction of bad ones would tend to about 1/3 with many tests, while if B is true it should tend to 1/6. Suppose that it is found to tend to the threshold value (4.24), close to 1/4. On sufficiently large m, you and I would then become skeptical about A and B; but intuition tells us that this would require a much larger m than ten, which was enough to make us and the robot skeptical when we find them all bad. Do the equations agree with our intuition here, if a new hypothesis F is introduced which specifies $P(\text{bad}|F\,X) \approx 1/4$?

Answer:

Let us devise hypotheses C and E, pertaining chances 99/100 and 1/100 of bad widgets. We thus have, for a certain hypothesis $H \in \{A,\, B,\, C,\, E\}$,

$$e(H|D\,X) = e(H|X) + 10\log_{10}\left[\frac{P(D|H\,X)}{\sum_{H' \neq H} w_{H'}^{(H)} P(D|H'\,X)}\right],$$

with

$$\begin{aligned} P(D|H\,X) &= [P(\text{bad}|H\,X)]^{m_b}\,[1 - P(\text{bad}|H\,X)]^{m-m_b} \\ &= \left(\left[\frac{P(\text{bad}|H\,X)}{1 - P(\text{bad}|H\,X)}\right]^{m_b/m} [1 - P(\text{bad}|H\,X)]\right)^m. \end{aligned}$$

It is then evident that the relevant comparison is between the terms in parentheses. Namely, we're concerned with

$$\begin{aligned} \left(\frac{P(D|H\,X)}{P(D|H'\,X)}\right)^{1/m} &= \left(\frac{P(\text{bad}|H\,X)}{P(\text{bad}|H'\,X)}\right)^{m_b/m} \left(\frac{P(\text{good}|H\,X)}{P(\text{good}|H'\,X)}\right)^{m_g/m} \\ &= \left(\frac{P(\text{bad}|H\,X)}{P(\text{bad}|H'\,X)}\frac{(1 - P(\text{bad}|H'\,X))}{(1 - P(\text{bad}|H\,X))}\right)^{m_b/m} \frac{1 - P(\text{bad}|H\,X)}{1 - P(\text{bad}|H'\,X)}, \end{aligned}$$

which gives the general threshold

$$f(H,\, H') = \frac{\log\left[\frac{P(\text{good}|H'\,X)}{P(\text{good}|H\,X)}\right]}{\log\left[\frac{P(\text{bad}|H\,X)}{P(\text{bad}|H'\,X)}\,\frac{P(\text{good}|H'\,X)}{P(\text{good}|H\,X)}\right]}.$$

In case the symmetry ain't evident, you can see that

$$f(H,\, H') = \frac{\log P(\text{good}|H'\,X) - \log P(\text{good}|H\,X)}{\log\left[\frac{P(\text{bad}|H\,X)}{P(\text{good}|H\,X)}\right] - \log\left[\frac{P(\text{bad}|H'\,X)}{P(\text{good}|H'\,X)}\right]}.$$

Intuition tells that this value is between the rates $f_H = P(\text{bad}|H\,X)$ and $f_{H'} = P(\text{bad}|H'\,X)$, which can be seen to be true as follows. For simplicity, suppose $P(\text{bad}|H\,X) > P(\text{bad}|H'\,X)$, and observe that the quotient $(P(D|H\,X)/P(D|H'\,X))^{1/m}$ is monotonously increasing on m_b/m. Further, we see that this threshold is already reached if $m_b/m = f_H$, since the function $x \mapsto x^{f_H}(1-x)^{1-f_H}$ has a maximum at

$$\begin{aligned} f_H x^{f_H-1}(1-x)^{1-f_H} + (1-f_H)x^{f_H}(1-x)^{-f_H} &= 0, \\ f_H\frac{1-x}{x} + (1-f_H) &= 0, \\ x &= f_H, \end{aligned}$$

so that

$$\begin{aligned} \left(\frac{f_H}{f_{H'}}\right)^{f_H}\left(\frac{1-f_H}{1-f_{H'}}\right)^{1-f_H} &> 1, \\ \left(\frac{f_H}{f_{H'}}\right)^{f_{H'}}\left(\frac{1-f_H}{1-f_{H'}}\right)^{1-f_{H'}} &< 1, \end{aligned}$$

and thus m is bounded between these two values.

This all demonstrates that the introduction of hypotheses C and E both makes us skeptical of A and B whenever bad widget rates go too high or drop too low, respectively.

Now, the critical value for hypothesis F, that $P(\text{bad}|F) = 1/4$, will be given by

$$\begin{aligned} P(F|D\,X) &= P(\overline{F}|D\,X), \\ P(D|F\,X)P(F|X) &= \sum_{H\neq F} P(D|H\,X)P(H|X), \\ \left(P(\text{bad}|F\,X)^{m_b/m}P(\text{good}|F\,X)^{m_g/m}\right)^m P(F|X) &= \sum_{H\neq F}\left(P(\text{bad}|H\,X)^{m_b/m}P(\text{good}|F\,X)^{m_g/m}\right)^m P(H|X). \end{aligned}$$

For big m this could be replaced with a pairwise comparison with the closest hypothesis, under a metric we don't know but probably differentiates more between values closer to $1/2$. So I'll lazily compare it with $1/3$:

$$\begin{aligned} m &= \frac{\log\left[\frac{P(A|X)}{P(F|X)}\right]}{\log\left[\left(\frac{P(\text{bad}|F\,X)}{P(\text{good}|A\,X)}\right)^{m_b/m}\left(\frac{P(\text{good}|F\,X)}{P(\text{good}|A\,X)}\right)^{m_g/m}\right]} \\ &= \frac{\log\left[\frac{P(A|X)}{P(F|X)}\right]}{\frac{m_b}{m}\log\left[\frac{P(\text{bad}|F\,X)}{P(\text{bad}|A\,X)}\right] + \frac{m_g}{m}\log\left[\frac{P(\text{good}|F\,X)}{P(\text{good}|A\,X)}\right]}. \end{aligned}$$

It is easily seen that the closest the predictions between hypotheses, the harder to differentiate them. The inverse of the denominator takes the value 140 (base 10). If we pick $P(F|X) \approx 1/100$ and $P(A|X) \approx 1/11$ we then get the final value of $m = 134$. The numerical solution involving hypotheses A, B, and F (script ""4.3.py") yields the value 2705... high.

Ex. 4.6:

I'm ommiting this one. It involves going back to chapters 3 and 4 and reviewing which problems can't be solved by application of a different set of rules,

$$\begin{aligned} p(\overline{A}) &= 1 - p(A), \\ p(A+B) &= p(A) + p(B) \text{ for mutually exclusive } A \text{ and } B, \\ p(AB) &= p(A)p(B) \text{ for "independent" } A \text{ and } B. \end{aligned}$$

It is easily seen that there's no definition of $A|B$, and I don't see, in a glance, how these can be used to resolve $p(AB)$ when AB are not independent.

Ex. 5.1:

Ommited so far. Involves assigning numerical degrees of belief in diverse affirmations. Will re-visit though.

Ex. 5.2: From these equations, find the exact conditions on (x, y, a, b) for divergence on the probability scale; that is,

$$|P(S|DI_X) - P(S|DI_Y)| > |P(S|I_X) - P(S|I_Y)|.$$

Answer:

The equations Edwin is refering to concern the proposition S that a certain drug is safe, with data D that a certain Mr. N claimed in T.V. that the drug is unsafe. Then, Mr. X and Mr. Y both have the same degree of trust in Mr. N:

$$\begin{aligned} P(D|SI_X) &= P(D|SI_Y) = a, \\ P(D|\overline{S}I_X) &= P(D|\overline{S}I_Y) = b. \end{aligned}$$

Here $a < b$ makes Mr. N more likely to be telling the truth. They however differ on their initial degree of trust in the drug's safety:

$$\begin{aligned} P(S|I_X) &= x, \\ P(S|I_Y) &= y, \end{aligned}$$

which finaly leads to

$$\begin{aligned} P(S|DI_X) &= \frac{ax}{ax + b(1-x)}, \\ P(S|DI_Y) &= \frac{ay}{ay + b(1-y)}. \end{aligned}$$

The condition for divergence is thus, in these terms:

$$\left| \frac{ax}{ax + b(1-x)} - \frac{ay}{ay + b(1-y)} \right| > |x - y|.$$

The special case $a = b$, where Mr. N is completely untrustworthy and conveys no information on his data, we get (as seen in the book), that the data changes nothing on Mr. X and Mr. Y's opinion. This equation reflects this by having no solution: rather than converging or diverging, they remain unchanged.

In the other cases, some mathsages lead to the expresion

$$ab > |(ax + b(1-x))(ay + b(1-y))|.$$

The value inside the absolute value is quadratic on x and y, and symmetric. We're thus dealing with a conic section symmetric about axes rotated 45 w.r.t. the original pair of cartesian axes. This can be resolved analitically, but

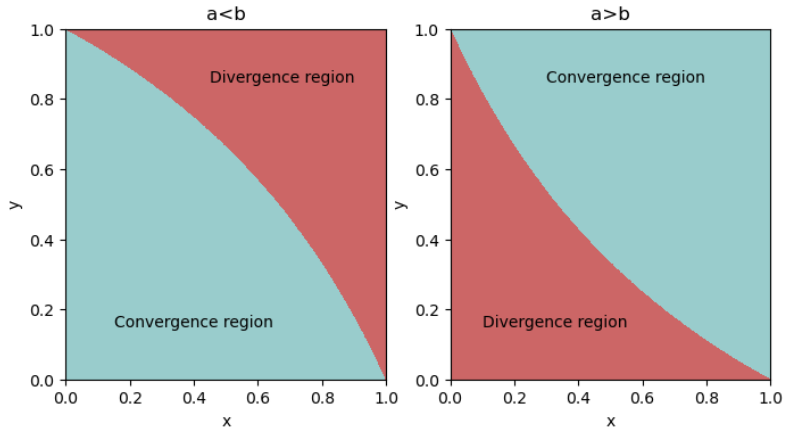

Figure 1: Regions of convergence and divergence for a liar Mr. N (left) and a not so liar Mr. N (right).

I'm not sure there's much purpose. Turns out to be four hyperbolas, with foci in $(0,0)$ and another point along the span$(1,1)$ axis which depends on a and b.

Figure **??** shows the convergence and divergence regions for $a < b$ (Mr. N being kinda trustworthy) and $a > b$, the ratio a/b controling the exentricity.

In the case where Mr. N is seen as trustworthy, they region of convergence is that of general agreement in that the drug is unsafe, their opinions diverging if they both believe the drug to be safe. The opposite happens whenever Mr. N is seen as a liar in general. They converge if they already kinda agreed that the pill was safe. I expected the divergence region to measure lack of agreement...

Ex. 5.3: It is evident from (5.31) that Mr. X and Mr. Y can never experience a reversal of viewpoint; that is, if initially Mr. X believes more strongly than Mr. Y in the safety of the drug, this will remain true whatever the values of a, b. Therefore, a necessary condition for reversal must be that they have different conditions about Mr. N; $a_x \neq a_y$ and/or $b_x \neq b_y$. But this does not prove that reversal is actually possible, so more analysis is needed. If reversal is possible, find a sufficient condition on (x, y, a_x, a_y, b_x, b_y) for this to take place, and illustrate it by a verbal scenario like the above. If it is not possible, prove this and explain the intuitive reason why reversal cannot happen.

Answer:

The equations for the log posterior for the drug being safe now take the form

$$\log\left[\frac{P(S|DI_X)}{P(\overline{S}|DI_X)}\right] = \log\left[\frac{x}{1-x}\right] + \log\left[\frac{a_X}{b_X}\right],$$
$$\log\left[\frac{P(S|DI_Y)}{P(\overline{S}|DI_Y)}\right] = \log\left[\frac{y}{1-y}\right] + \log\left[\frac{a_Y}{b_Y}\right],$$

so that their difference is

$$\log\left[\frac{x}{1-x}\right] - \log\left[\frac{y}{1-y}\right] + \log\left[\frac{a_X}{b_X}\right] - \log\left[\frac{a_Y}{b_Y}\right].$$

So, inversion of beliefs is acchieved by requiring that the difference of log likelihoods be greater in magnitude than the difference of original beliefs, with opposite sign. For the sake of concreteness, if $x = 0.6$ and $y = 0.4$, it suffices to take $a_Y/b_Y > 0.6$ and $a_X/b_X < 0.4$. Say, $a_Y = b_X = 0.8$, $b_Y = a_X = 0.2$; this takes an original degree of belief from (log prior) 0.81 to -1.96.

In this example, Mr. X's belief is that the drug is most probably safe, and Mr. Y's is the opposite. But Mr. X is so convinced that Mr. N wouldn't lie, and Mr. Y is so convinced of the contrary, that their stances after this encounter become reversed. Cool.

Ex. 5.4: Our story has a curious sequel. In turn, it was noticed that Neptune was not following exactly its proper course, and so one naturally assumed that there is still another planet causing this. Percivall Lowel, by a

similar calculation, predicted its orbit, and Clyde Tombaugh proceeded to find the new planet (Pluto), although not so close to the predicted position. But now the story changes: modern data on the motion of Pluto's moon indicated that the mass of Pluto is too small to have caused the perturbation of Neptune which motivated Lowell's calculation. Thus, the discrepancies in the motions of Neptune and Pluto were unaccounted for (We are indebted to Dr. Brad Schaeter for this information). Try to extend our probability analysis to take this new circumstance into account; at this point, where did Newton's theory stand? For more background information, see Hoyt (1980) or Whyte (1980). More recently, it appears that the mass of Pluto had been estimated wrongly and the discrepancies were after all not reall; then it seems that the status of Newton's theory should revert to its former one. Discuss this sequence of pieces of information in terms of probability theory. Do we update by Bayes' theorem as each new fact comes in? Or do we just return to the beginning when we learn that a previous datum was false?