Additional Problem Set for Lecture 5: Confirmation

(1) Let P be a probability distribution defined over the binary variables H (with values H and $\neg H$) and E (with values E and $\neg E$). Furthermore, $P(\mathrm{H}, \mathrm{E}) = 0.4, P(\mathrm{H}, \neg\mathrm{E}) = 0.2$ and $P(\neg\mathrm{H}, \mathrm{E}) = 0.1$. (a) What is $P(\neg\mathrm{H}, \neg\mathrm{E})$? (b) What are $P(\mathrm{H})$ and $P(\mathrm{E})$? (c) What is $P(\mathrm{H}|\mathrm{E})$? (d) Does E confirm H?

(2) A confirmation measure $c(\mathrm{H}, \mathrm{E})$ measures how strongly a piece of evidence E confirms a hypothesis H. We assume that $P(\mathrm{H}) \in (0, 1)$, i.e. that $0 < P(\mathrm{H}) < 1$. E confirms H if $c > 0$, E disconfirms H if $c < 0$, and E is irrelevant for H if $c = 0$. Here are three popular confirmation measures: (i) the *distance measure* $d(\mathrm{H}, \mathrm{E}) := P(\mathrm{H}|\mathrm{E}) - P(\mathrm{H})$, the *log-ratio measure* $r(\mathrm{H}, \mathrm{E}) := \log\left(P(\mathrm{H}|\mathrm{E})/P(\mathrm{H})\right)$, and the *log-likelihood measure* $l(\mathrm{H}, \mathrm{E}) := \log\left(P(\mathrm{E}|\mathrm{H})/P(\mathrm{E}|\neg\mathrm{H})\right)$. N.B.: The bases of the logarithms should be positive. As a convention, let us choose the basis e, i.e. the natural logarithm. (a) Show that these measures satisfy the above-mentioned properties. (b) Work out the values of $d(\mathrm{H}, \mathrm{E}), r(\mathrm{H}, \mathrm{E})$ and $l(\mathrm{H}, \mathrm{E})$ for the probability model in problem (1). (c) Interestingly, the measures d, r and l are not ordinally equivalent, i.e. they do not always imply the same ordering of hypotheses H_1, $\mathrm{H}_2, \ldots$ (given some evidence E). Read Branden Fitelson's classic paper "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity" (http://fitelson.org/psa.pdf) to learn more about this issue. Can you come up with your own example?

(3) There is a new test for a rare (and, indeed, rather horrible) disease. Only 2% of the patients who are tested for the disease actually have it. Unfortunately, like all medical tests, the new test is not fully reliable. 5% of the tests report that the patient has the disease although she does not have it ("false positives"). And 1% of the tests report that the patient does not have the disease although she does has it ("false negatives"). (a) A patient is tested and the result is positive. What is the probability that she has the disease? (b) The patient is tested negative. What is the probability that she has the disease?

(4) A university has two departments, D1 and D2. This year, 100 new students applied, of which 70 were male and 30 female. 65 male students applied to D1 of which 45 were accepted. 10 female students applied to D1, of which 9 were accepted. 5 male students applied to D2 of which 2 were accepted. 20 female students applied to D2, of which 9 were accepted. (a) What is the acceptance rate of male and female students in D1 and

in D2? (b) What is the acceptance rate of male and female students in the university (i.e. in both departments, taken together)? Do you find the results surprising? Give an informal explanation.

(5) Let us assume that E confirms the hypothesis H. (a) Show that then H confirms E. (b) There is another hypothesis, H', and we assume that E is conditionally independent of H' given H, i.e. it holds that $P(\mathrm{E}|\mathrm{H},\mathrm{H}') = P(\mathrm{E}|\mathrm{H})$. Show that E also confirms H $\wedge$ H'. Do you find this result intuitive?

(6) Evidence E_1 is conditionally independent of evidence E_2 given the hypothesis H (in symbols: $E_1 \perp\!\!\!\perp E_2 | H$) if $P(\mathrm{E}_1|\mathrm{E}_2,\mathrm{H}) = P(\mathrm{E}_1|\mathrm{H})$ and accordingly for the other instantiations of the three propositional variables. That is, learning E_2 does not change the probability of E_1 if H is known. The truth or falsity of E_1 only depends on H. (a) Show that this condition is equivalent to $P(\mathrm{E}_1,\mathrm{E}_2|\mathrm{H}) = P(\mathrm{E}_1|\mathrm{H})\,P(\mathrm{E}_2|\mathrm{H})$. (b) Similarly, show that $P(\mathrm{E}_1|\mathrm{E}_2,\neg\mathrm{H}) = P(\mathrm{E}_1|\neg\mathrm{H})$ is equivalent to $P(\mathrm{E}_1,\mathrm{E}_2|\neg\mathrm{H}) = P(\mathrm{E}_1|\neg\mathrm{H})\,P(\mathrm{E}_2|\neg\mathrm{H})$. (c) Calculate the posterior probability of a hypothesis H after learning two conditionally independent pieces of evidence E_1 and E_2 (i.e. for which $E_1 \perp\!\!\!\perp E_2 | H$ holds). Assume that both pieces of evidence have the same likelihoods, i.e. $P(\mathrm{E}_1|\mathrm{H}) = P(\mathrm{E}_2|\mathrm{H}) =: p$ and $P(\mathrm{E}_1|\neg\mathrm{H}) = P(\mathrm{E}_2|\neg\mathrm{H}) =: q$ and that each piece of evidence confirms H, i.e. that $p > q$. (d) Let $p = 0.8$ and $q = 0.2$ and compare the confirmation, using the difference measure d (see problem 2), for two (conditionally independent) pieces of evidence with the confirmation which one gets from one piece of evidence. (e) Generalize the results to n conditionally independent pieces of evidence $E_1, \ldots, E_n$ with the same likelihoods p and q.

(7) Let P be a probability distribution over the binary propositional variables E and H with $P(\mathrm{H}) \in (0,1)$ and $P(\mathrm{E}) = 1$. (a) Show that $P(\mathrm{E}|\mathrm{H}) = P(\mathrm{E}|\neg\mathrm{H}) = 1$. (b) Show that, $d(\mathrm{H},\mathrm{E}) = r(\mathrm{H},\mathrm{E}) = l(\mathrm{H},\mathrm{E}) = 0$, i.e. old evidence does not confirm. Do you find this result intuitive?

Solutions

(1a) $P(\mathrm{H},\mathrm{E}) + P(\mathrm{H},\neg\mathrm{E}) + P(\neg\mathrm{H},\mathrm{E}) + P(\neg\mathrm{H},\neg\mathrm{E}) = 1$. Hence, $P(\neg\mathrm{H},\neg\mathrm{E}) = 0.3$.

(1b) $P(\mathrm{H}) = P(\mathrm{H},\mathrm{E}) + P(\mathrm{H},\neg\mathrm{E}) = 0.6$ and $P(\mathrm{E}) = P(\mathrm{H},\mathrm{E}) + P(\neg\mathrm{H},\mathrm{E}) = 0.5$.

(1c) $P(\mathrm{H}|\mathrm{E}) = P(\mathrm{H},\mathrm{E})/P(\mathrm{E}) = (0.4)/(0.5) = 0.8$.

(1d) Yes, E confirms H because $P(\mathrm{H}|\mathrm{E}) = 0.8 > P(\mathrm{H}) = 0.6$.

(2a) Here we only provide three hints. First, note that E confirms H iff $P(\mathrm{H}|\mathrm{E}) > P(\mathrm{H})$, i.e. iff $P(\mathrm{H}|\mathrm{E}) - P(\mathrm{H}) > 0$, i.e. iff $P(\mathrm{H}|\mathrm{E})/P(\mathrm{H}) > 1$. Second, observe that $f(x) = \log x$ is a strictly monotonically increasing in x and that $\log 1 = 0$. See http://en.wikipedia.org/wiki/Logarithm. Third, note our discussion of the likelihood ratio in the above quiz.

(2b) $d(\mathrm{H},\mathrm{E}) = 0.8 - 0.6 = 0.2$ and $r(\mathrm{H},\mathrm{E}) = \log((0.8)/(0.6)) = \log(4/3) \approx 0.288$. Next we calculate $P(\mathrm{E}|\mathrm{H}) = P(\mathrm{H},\mathrm{E})/P(\mathrm{H}) = (0.4)/(0.6) = 2/3$ and $P(\mathrm{E}|\neg\mathrm{H}) = P(\neg\mathrm{H},\mathrm{E})/P(\neg\mathrm{H}) = (0.1)/(0.4) = 1/4$. Hence, $l(\mathrm{H},\mathrm{E}) = \log(8/3) \approx 0.981$.

(3) Let us introduce two binary propositional variables, H and E. H has the values H: The patient has the disease, and ¬ H: The patient does not have the disease. E has the values E: The test was positive, and ¬ E: The test was negative. With this, we can fix $P(\mathrm{H}) = 0.02, P(\mathrm{E}|\neg\mathrm{H}) = 0.05$ and $P(\neg\mathrm{E}|\mathrm{H}) = 0.01$. Hence, $P(\mathrm{E}|\mathrm{H}) = 1 - P(\neg\mathrm{E}|\mathrm{H}) = 0.99$. (3a) Inserting everything into Bayes Theorem, we obtain

$$\begin{aligned} P(\mathrm{H}|\mathrm{E}) &= \frac{P(\mathrm{E}|\mathrm{H})\,P(\mathrm{H})}{P(\mathrm{E}|\mathrm{H})\,P(\mathrm{H}) + P(\mathrm{E}|\neg\mathrm{H})\,P(\neg\mathrm{H})} \\ &= \frac{0.99 \times 0.02}{0.99 \times 0.02 + 0.05 \times 0.98} \approx 0.288. \end{aligned}$$

(3b) Similarly, we obtain

$$\begin{aligned} P(\mathrm{H}|\neg\mathrm{E}) &= \frac{P(\neg\mathrm{E}|\mathrm{H})\,P(\mathrm{H})}{P(\neg\mathrm{E}|\mathrm{H})\,P(\mathrm{H}) + P(\neg\mathrm{E}|\neg\mathrm{H})\,P(\neg\mathrm{H})} \\ &= \frac{0.01 \times 0.02}{0.01 \times 0.02 + 0.95 \times 0.98} \approx 0.0002. \end{aligned}$$

See also Figures 1 and 2 for illustration. (N.B.: The plots were made using the software *Mathematica*.)

(4a) The acceptance rate of male students in D1 is $45/65 \approx 0.69$. The acceptance rate of female students in D1 is $9/10 = 0.90$. That is, the acceptance rate of female students is higher than the acceptance rate for male students in D1. The acceptance rate of male students in D2 is $2/5 = 0.40$. The acceptance rate of female students in D2 is $9/30 = 0.45$.

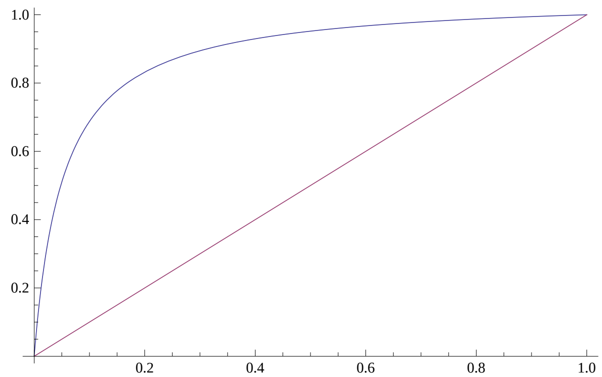

Figure 1: $P(\mathrm{H}|\mathrm{E})$ with the rates of false positives and false negatives specified in problem 3 as a function of the prior probability $P(\mathrm{H})$. The effect of the test is the excess of the probability compared to the red line (no test).

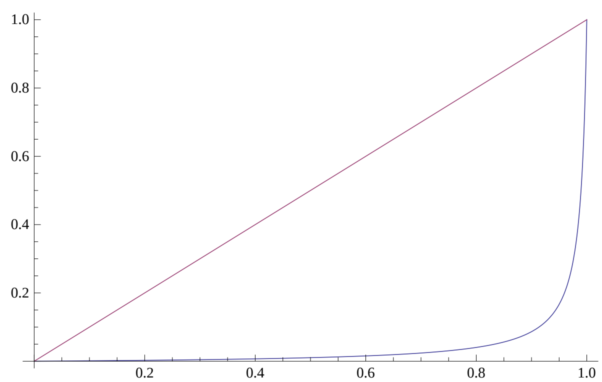

Figure 2: $P(\mathrm{H}|\neg\mathrm{E})$ with the rates of false positives and false negatives specified in problem 3 as a function of the prior probability $P(\mathrm{H})$. The effect of the test is the excess of the probability compared to the red line (no test).

That is, the acceptance rate of female students is higher than the acceptance rate for male students in D2.

(4b) The university accepts $45 + 2 = 47$ male students, i.e. the acceptance rate of male students is $47/70 \approx 0.67$. The university accepts $9 + 9 = 18$ female students, i.e. the acceptance rate of female students is $18/30 = 0.60$. That is, the acceptance rate of female students is *lower* than the acceptance rate for male students in the whole university. This counterintuitive example is an instance of *Simpson's Paradox*. Read more about it in the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/paradox-simpson/).

(5a) E confirms H iff $P(\mathrm{H}|\mathrm{E}) > P(\mathrm{H})$, i.e. iff $P(\mathrm{H}|\mathrm{E})/P(\mathrm{H}) > 1$. Using Bayes Rule, we obtain: H confirms E iff $(P(\mathrm{E}|\mathrm{H})\, P(\mathrm{H}))/(P(\mathrm{H})\, P(\mathrm{E})) = P(\mathrm{E}|\mathrm{H})/P(\mathrm{E}) > 1$, i.e. iff H confirms E.

(5b) We use the definition of conditional probability twice and write

$$\begin{aligned} P(\mathrm{H}, \mathrm{H}'|\mathrm{E}) &= \frac{P(\mathrm{H}, \mathrm{H}', \mathrm{E})}{P(\mathrm{E})} \\ &= \frac{P(\mathrm{E}|\mathrm{H}, \mathrm{H}')\, P(\mathrm{H}, \mathrm{H}')}{P(\mathrm{E})}. \end{aligned}$$

Next, we use the independence condition, i.e. that $P(\mathrm{E}|\mathrm{H}, \mathrm{H}') = P(\mathrm{E}|\mathrm{H})$. Hence,

$$P(\mathrm{H}, \mathrm{H}'|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H})}{P(\mathrm{E})} \cdot P(\mathrm{H}, \mathrm{H}')$$

As E confirms H, we conclude from (5a) that $P(\mathrm{E}|\mathrm{H})/P(\mathrm{E}) > 1$. Hence, $P(\mathrm{H}, \mathrm{H}'|\mathrm{E}) > P(\mathrm{H}, \mathrm{H}')$, which means that E confirms H $\wedge$ H'. This is surprising as H' is, according to the assumption, independent of E (given H). And set, it is part of a conjunct which is confirmed. This problem is an instance of the so-called problem of irrelevant conjunction. There are a number of excellent papers on the topic which you can download from the web. Go ahead if you are interested!

(6a) We use the definition of conditional probability and obtain the following sequence of equivalent equations:

$$\begin{aligned}
P(\mathrm{E}_1|\mathrm{E}_2,\mathrm{H}) &= P(\mathrm{E}_1|\mathrm{H}) \\
\frac{P(\mathrm{E}_1,\mathrm{E}_2,\mathrm{H})}{P(\mathrm{E}_2,\mathrm{H})} &= P(\mathrm{E}_1|\mathrm{H}) \\
\frac{P(\mathrm{E}_1,\mathrm{E}_2,\mathrm{H})}{P(\mathrm{E}_2|\mathrm{H})\,P(\mathrm{H})} &= P(\mathrm{E}_1|\mathrm{H}) \\
\frac{P(\mathrm{E}_1,\mathrm{E}_2,\mathrm{H})}{P(\mathrm{H})} &= P(\mathrm{E}_1|\mathrm{H})\,P(\mathrm{E}_2|\mathrm{H}) \\
P(\mathrm{E}_1,\mathrm{E}_2|\mathrm{H}) &= P(\mathrm{E}_1|\mathrm{H})\,P(\mathrm{E}_2|\mathrm{H})
\end{aligned}$$

(6b) Accordingly.

(6c) Using Bayes Theorem and the rule of total probability, we obtain

$$P(\mathrm{H}|\mathrm{E}_1,\mathrm{E}_2) = \frac{P(\mathrm{E}_1,\mathrm{E}_2|\mathrm{H})\,P(\mathrm{H})}{P(\mathrm{E}_1,\mathrm{E}_2|\mathrm{H})\,P(\mathrm{H}) + P(\mathrm{E}_1,\mathrm{E}_2|\neg\mathrm{H})\,P(\neg\mathrm{H})}.$$

Next, we use the conditional independence of E_1 and E_2 given H:

$$P(\mathrm{H}|\mathrm{E}_1,\mathrm{E}_2) = \frac{P(\mathrm{E}_1|\mathrm{H})\,P(\mathrm{E}_2|\mathrm{H})\,P(\mathrm{H})}{P(\mathrm{E}_1|\mathrm{H})\,P(\mathrm{E}_2|\mathrm{H})\,P(\mathrm{H}) + P(\mathrm{E}_1|\neg\mathrm{H})\,P(\mathrm{E}_2|\neg\mathrm{H})\,P(\neg\mathrm{H})}$$

To simplify notation, we set $P(\mathrm{H}) := h$ and $P(\neg\mathrm{H}) = 1 - h =: \bar{h}$ and obtain after some algebra:

$$P(\mathrm{H}|\mathrm{E}_1,\mathrm{E}_2) = \frac{h}{h + \bar{h}\,(q/p)^2} \tag{1}$$

(6d) For one piece of evidence, one obtains:

$$P(\mathrm{H}|\mathrm{E}_1) = \frac{h}{h + \bar{h}\,(q/p)}$$

For the values of p and q given above, we obtain:

$$\begin{aligned}
P(\mathrm{H}|\mathrm{E}_1) &= \frac{h}{h + \bar{h}\,(1/4)} = \frac{4h}{3h+1} \\
P(\mathrm{H}|\mathrm{E}_1,\mathrm{E}_2) &= \frac{h}{h + \bar{h}\,(1/4)^2} = \frac{16h}{15h+1}
\end{aligned}$$

We can now calculate the difference measure d for both scenarios:

$$\begin{aligned}
d(\mathrm{H},\mathrm{E}_1) &= P(\mathrm{H}|\mathrm{E}_1) - P(\mathrm{H}) = \frac{3h\bar{h}}{3h+1} \\
d(\mathrm{H},\mathrm{E}_1,\mathrm{E}_2) &= P(\mathrm{H}|\mathrm{E}_1,\mathrm{E}_2) - P(\mathrm{H}) = \frac{15h\bar{h}}{15h+1}
\end{aligned}$$

Figure 3 shows both expressions as a function of the prior probability h. Clearly, collecting the additional piece of evidence E_2 increases the confirmation of H. Explain the shape of the curves!

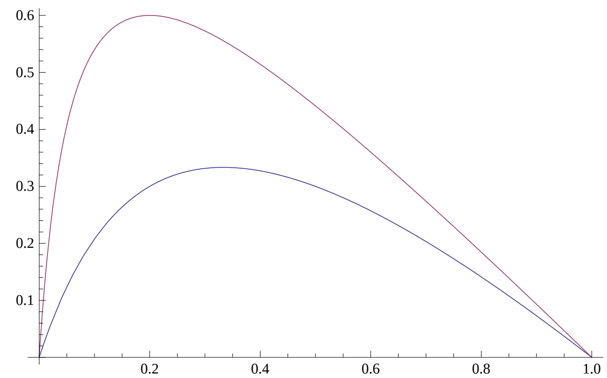

Figure 3: $d(\mathrm{H}, \mathrm{E}_1, \mathrm{E}_2)$ (red) and $d(\mathrm{H}, \mathrm{E}_1)$ (blue) as a function of the prior probability h.

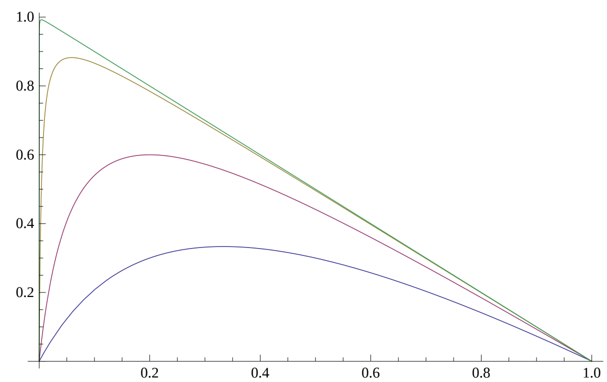

Figure 4: The difference measure for 1, 2, 4, and 8 pieces of evidence (from bottom to top) as a function of the prior probability h.

(6e) Eq. (1) can be generalized straight-forwardly to n pieces of evidence (with the same likelihoods as before):

$$P(\mathrm{H}|\mathrm{E}_1, \ldots, \mathrm{E}_\mathrm{n}) = \frac{h}{h + \bar{h}\,(q/p)^n}$$

Figure 4 shows the corresponding difference measures for 1, 2, 4, and 8 pieces of evidence as a function of the prior probability h.

(7a) We apply the rule of total probability and write

$$P(\mathrm{E}) = P(\mathrm{E}|\mathrm{H})P(\mathrm{H}) + P(\mathrm{E}|\neg\mathrm{H})P(\neg\mathrm{H}) = 1$$

As $0 < P(\mathrm{H}), P(\neg\mathrm{H}) < 1$ and $P(\mathrm{H}) + P(\neg\mathrm{H}) = 1$, we conclude that $P(\mathrm{E}|\mathrm{H}) = P(\mathrm{E}|\neg\mathrm{H}) = 1$.

(7b) We first calculate the posterior probability

$$P(\mathrm{H}|\mathrm{E}) = \frac{P(\mathrm{E}|\mathrm{H})}{P(\mathrm{E})} \cdot P(\mathrm{H}) = \frac{1}{1} \cdot P(\mathrm{H}) = P(\mathrm{H}).$$

Hence, $d(\mathrm{H},\mathrm{E}) = P(\mathrm{H}|\mathrm{E}) - P(\mathrm{H}) = 0$. Furthermore, $r(\mathrm{H},\mathrm{E}) = \log\left(P(\mathrm{H}|\mathrm{E})/P(\mathrm{H})\right) = \log 1 = 0$. Finally, $l(\mathrm{H},\mathrm{E}) = \log\left(P(\mathrm{E}|\mathrm{H})/P(\mathrm{E}|\neg\mathrm{H})\right) = \log 1 = 0$.

Our calculation shows that evidence that is already known cannot confirm a hypothesis. Only evidence that is unexpected before the test is conducted can confirm. This apparently clashes with the practice of science, as Clark Glymour has argued. Read more about the so-called **problem of old evidence**, e.g. in the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/epistemology-bayesian/). I also recommend Glymour's classical chapter "Why I am not a Bayesian" (http://fitelson.org/probability/glymour.pdf). Here is an extra problem (for the experts): Work out $d(\mathrm{H},\mathrm{E}), r(\mathrm{H},\mathrm{E})$ and $l(\mathrm{H},\mathrm{E})$ if $P(\mathrm{E}) = 1 - \epsilon$ where ϵ is understood to be small. Does the problem of old evidence disappear if we are not fully certain about the evidence?