HW3, Theory of Inference 2015/6

Jonathan Rougier
School of Mathematics
University of Bristol UK

Here is a homework about Decision Theory, concerning a real and important decision. You are *strongly encouraged* to do this homework.

The London VAAC (`http://www.metoffice.gov.uk/aviation/vaac/`) assesses the ash concentration in the eastern N. Atlantic. We will consider one pixel, say the one at $(54, -31)$–$(55, -30)$. This pixel is either ashy ($X = \mathsf{true}$) or not-ashy ($X = \mathsf{false}$). A satellite retrieval provides an imperfect test of whether the pixel is ashy. This test is either positive ($Y = +$) or negative ($Y = -$), with $+$ favouring the presence of ash.

1. Define the 'sensitivity' and the 'specificity' of the retrieval; denote these as α and β in what follows. Propose a reason why the sensitivity might be high, and the specificity low. [5 marks]

Answer.

$$\begin{aligned}
\alpha &:= \Pr(Y = + \mid X = \mathsf{true}) && \text{(sensitivity)} \\
\beta &:= \Pr(Y = - \mid X = \mathsf{false}) && \text{(specificity).}
\end{aligned}$$

The sensitivity is usually high, but the specificity is low, because the test will often report $+$ in the presence of clouds, regardless of whether or not there is ash. These are termed 'false positives'. A more detailed analysis would include the presence of clouds.

2. What additional information do we need in order to compute the posterior probability of $X = \mathsf{true}$ given a positive test? (Clue: denote this quantity as π_+.) Give the formulae for the posterior probabilities of X conditional on $Y = +$. (You may want to express these as odds.) [10 marks]

   Answer. We also need the base rate, $\pi_+ := \Pr(X = \mathsf{true})$.[1] For meteorology, this is nearly zero if there has not been an ashy eruption, but much higher if there has. Then, by Bayes's Theorem,

$$\begin{aligned} O(+) &= \frac{\Pr(Y = + \mid X = \mathsf{true}) \cdot \Pr(X = \mathsf{true})}{\Pr(Y = + \mid X = \mathsf{false}) \cdot \Pr(X = \mathsf{false})} \\ &= \frac{\alpha}{1-\beta} \cdot \frac{\pi_+}{1-\pi_+}, \end{aligned}$$

   where $O(+)$ are the posterior odds for $Y = +$. Then

$$\begin{aligned} \Pr(X = \mathsf{true} \mid Y = +) &= \frac{O(+)}{1 + O(+)} \\ \Pr(X = \mathsf{false} \mid Y = +) &= \frac{1}{1 + O(+)}. \end{aligned}$$

3. Let the action set be $\mathcal{A} := \{\mathsf{safe}, \mathsf{unsafe}\}$. Denote the loss function as

$L(a, x)$	$x = \mathsf{false}$	$x = \mathsf{true}$
$a = \mathsf{safe}$	ℓ_{00}	ℓ_{01}
$a = \mathsf{unsafe}$	ℓ_{10}	ℓ_{11}

   What are the natural constraints on the values of the four ℓ's? [5 marks]

   Answer. We must have $\ell_{00} < \ell_{10}$ and $\ell_{11} < \ell_{01}$, on the grounds that wrong decisions incur larger losses than right ones. It would be quite common to have $\ell_{00} = \ell_{11}$. The worst possible decision is likely to be $(\mathsf{safe}, \mathsf{true})$, and so we would expect $\ell_{01} > \ell_{10}$.

[1] Apologies: it would have been better to have called this π_{true} but I changed notation and forgot to update this.

4. State the Bayes Rule Theorem. Prove that the Bayes Rule for this problem with $y = +$ is

$$\delta^*(+) = \begin{cases} \mathsf{safe} & O(+) < \frac{\ell_{10} - \ell_{00}}{\ell_{01} - \ell_{11}} \\ \mathsf{unsafe} & \text{otherwise,} \end{cases} \tag{1}$$

where

$$O(+) := \frac{\Pr(X = \mathsf{true} \mid Y = +)}{\Pr(X = \mathsf{false} \mid Y = +)},$$

termed the *posterior odds* for X when $Y = +$. [15 marks]

Answer. A decision rule is a function $\delta : y \mapsto a$. Let $\mathcal{D}$ be the set of all decision rules. Then the Bayes Rule δ^* satisfies

$$\delta^* = \operatorname*{argmin}_{\delta \in \mathcal{D}} \mathrm{E}\{L(\delta(Y), X)\}.$$

The Bayes Rule Theorem states that

$$\delta^*(y) = \operatorname*{argmin}_{a \in \mathcal{A}} \mathrm{E}\{L(a, X) \mid Y = y\} \qquad \forall y \in \mathcal{Y}.$$

In our case we want $\delta^*(+)$. Considering the two possibilities for a, we compute

$$\begin{aligned} \mathrm{E}\{L(\mathsf{safe}, X) \mid Y = +\} &= \ell_{00} \, \frac{1}{1 + O(+)} + \ell_{01} \, \frac{O(+)}{1 + O(+)} \\ &\propto \ell_{00} + \ell_{01} \, O(+) \\ \mathrm{E}\{L(\mathsf{unsafe}, X) \mid Y = +\} &= \ell_{10} \, \frac{1}{1 + O(+)} + \ell_{11} \, \frac{O(+)}{1 + O(+)} \\ &\propto \ell_{10} + \ell_{11} \, O(+) \end{aligned}$$

where the positive constant $1/(1 + O(+))$ can be dropped because we are comparing sizes. So we will choose safe for the Bayes Rule if and only if

$$\ell_{00} + \ell_{01} \, O(+) < \ell_{10} + \ell_{11} \, O(+),$$

or, on re-arranging,

$$O(+) < \frac{\ell_{10} - \ell_{00}}{\ell_{01} - \ell_{11}},$$

where the righthand side is positive.

5. Interpret this result. [5 marks]

Answer. This result is clearly oriented in the right direction, with the action switching from safe to unsafe as the posterior odds increases. We can normalise the losses by setting $\ell_{00} = \ell_{11} = 0$, without loss of generality. The ratio of the two 'wrong' losses sets the threshold. If, as we suggested, $\ell_{01} > \ell_{10}$, then the threshold is < 1, i.e. we switch from safe to unsafe before $X = \mathsf{true}$ becomes the most probable outcome. This is because we want to avoid the large loss in $(\mathsf{safe}, \mathsf{true})$.

6. The test shows $Y = +$. Using the values $\alpha = 0.95$, $\beta = 0.3$, $\pi_+ = 0.05$, $\ell_{00} = \ell_{11} = 0$, $\ell_{01}/\ell_{10} = 10$, what is the optimal action? Comment on this result. [5 marks]

Answer. We have

$$O(+) = \frac{0.95}{1 - 0.3} \cdot \frac{0.05}{1 - 0.05} = \frac{0.05}{0.7} \approx 0.07.$$

But $\ell_{10}/\ell_{01} = 0.1$, and hence the optimal action is $\delta^*(+) = \mathsf{safe}$. What is interesting about this is that the optimal action is safe even though the test is positive. We infer that the optimal action is also safe when the test is negative. Ie, the meteorologists at the VAAC can all go home. Actually, they need to figure out how to improve their test. One problem is the high rate of false positives (low specificity). If they could incorporate information about clouds, which allowed them to be distinguished, partly, from ash, then they could derive a better decision rule, with a lower expected loss.