COMS10003 Work Sheet 12

Probability II: Bayes' Theorem

Andrew Calway December 16, 2014

This worksheet is about Bayes' Theorem. It is one of those topics in which lots of practice is pretty much essential. But when you've got it you'll be well pleased, bit like riding a bike. As before, some questions have been taken, or adapted, from my bookshelf books:

Probability, Random Variables and Stochastic Processes by A.Papoulis, McGraw-Hill
Linear Algebra and Probability for Computer Science Applications by E.Davis, CRC Press

1. A quick starter for ten: if two events are mutually exclusive, can they be independent?

   Answer: *No, they can't. If two events are ME, then they can't exist at the same time, i.e. one excludes the other, and thus the occurrence of one does indeed impact on the possible occurrence of the other. So they can't be independent.*

2. A patient has a test for a disease which affects 1 in 10,000 people. If the patient has the disease then with 99% probability the test will be positive; if the patient doesn't have the disease then with 99% probability the test will be negative. If the test is positive, what is the probability that the patient has the disease?

   Answer: *Need to compute $P(D|T)$. From Bayes'*

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\neg D)P(\neg D)} \\ &= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \\ &= 0.98\% \end{aligned}$$

3. We have four boxes. Box 1 contains 2000 components of which 5% are defective. Box 2 contains 500 components of which 40% are defective. Boxes 3 and 4 each contain 1000 components and 10% of each are defective. We select at random one of the boxes and remove at random one component.

   (a) What is the probaility that selected component is defective?

   (b) If we find that the component is defective, what is the probability that it came from box 2?

   Answer: *(a)* $P(defective) = 0.25 * (0.05 + 0.4 + 0.1 + 0.1) = 0.1625$
   (b) $P(B2|def) = \frac{P(def|B2)P(B2)}{P(def)} = \frac{0.4 \times 0.25}{0.1625} = 0.615(62\%)$

4. Box 1 contains 1 white ball and 99 red balls. Box 2 contains 1 red ball and 99 white balls. A ball is picked from a randomly selected box. If the ball is red what is the probability that it came from box 1.

   Answer: $P(B1|R) = \frac{P(R|B1)P(B1)}{P(R|B1)P(B1)+P(R|B2)P(B2)} = \frac{0.99\times 0.5}{0.5(0.99+0.01)} = 0.99(99\%)$

5. Box 1 contains 1000 bulbs of which 10% are defective. Box 2 contains 2000 bulbs of which 5% are defective. Two bulbs are picked from a randomly selected box.

   (a) Find the probability that both are defective.

   (b) Assuming both are defective, find the probability that they came from box 1.

   Answer: *(a)* $P(bothdef) = 0.5(0.1 \times 99/999 + 0.05 \times 99/1999) = 0.0062(0.62\%)$

   (b) $P(B1|bothdef) = \frac{P(bothdef|B1)P(B1)}{P(bothdef)} = \frac{(0.1\times 99/999)\times 0.5}{0.0062} = 0.8(80\%)$

6. Over a period of one week, 2000 emails were collected. Of these, 1500 were spam and 500 were not spam. The presence or not of three words w_1, w_2 and w_3 were then noted in each email. The results are shown below.

word	# spam emails containing *word*	# non-spam emails containing *word*
w_1	1000	100
w_2	200	300
w_3	500	200

   (a) If a new email contains all three words, is it more likely to be spam or non-spam?

   (b) What if it contains words w_1 and w_3 but not w_2?

   Answer: *(a)*

   $P(spam) = 3/4 \quad P(\neg spam) = 1/4$

$P(w_1\|spam)$	*2/3*		$P(w_1\|\neg spam)$	*1/5*
$P(w_2\|spam)$	*2/15*		$P(w_2\|\neg spam)$	*3/5*
$P(\neg w_2\|spam)$	*13/15*		$P(\neg w_2\|\neg spam)$	*2/5*
$P(w_3\|spam)$	*1/3*		$P(w_3\|\neg spam)$	*2/5*

   $\frac{P(spam|w_1,w_2,w_3)}{P(\neg spam|w_1,w_2,w_3)} = \frac{(2/3)(2/15)(1/3)(3/4)}{(1/5)(3/5)(2/5)(1/4)} = 1.85$

   Therefore almost twice as likely to be spam

   (b)

   $\frac{P(spam|w_1,\neg w_2,w_3)}{P(\neg spam|w_1,\neg w_2,w_3)} = \frac{(2/3)(13/15)(1/3)(3/4)}{(1/5)(2/5)(2/5)(1/4)} = 18.06$

   Therefore 18 times more likely to be spam.