Problem Sheet for Chapter 1

1. Consider a standard pack of 52 playing cards. You pick a card at random, what is the probability you pick:
   (a) A Queen, given you have picked a picture card (King, Queen, Jack)?
   (b) The five of clubs, given you have picked a black card?
   (c) A black card, given you have not picked the five of clubs?

2. Decide if each of the following events can be interpreted as happening with probability $p \in (0, 1)$ by frequentists:
   (a) The Bermuda triangle exists.
   (b) Getting a 6 when rolling a dice.
   (c) Someone will test positive for Covid-19 after contracting the disease.

3. An urn contains three coins. Two of the coins are fair, but one of the coins has heads on both sides.
   (a) You pick a coin out of the urn without looking and flip it. What's the probability you get heads?
   (b) You pick a coin out of the urn without looking and flip it and get heads. What's the probability it's the two-headed coin?

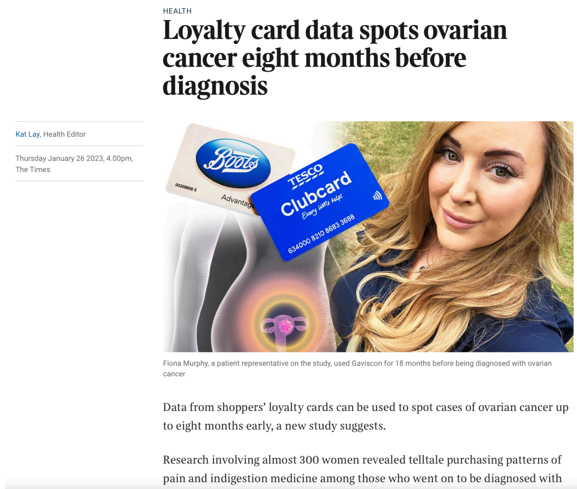
4. You see a sponsored post online containing the word *bitcoin*. You want to work out the probability the post is spam.
   (a) Using the law of total probability, show the probability that the post is spam, given it contains the word bitcoin is

$$\pi(\text{spam} \mid \text{bitcoin}) = \frac{\pi(\text{bitcoin} \mid \text{spam})\pi(\text{spam})}{\pi(\text{bitcoin} \mid \text{spam})\pi(\text{spam}) + \pi(\text{bitcoin} \mid \text{not spam})\pi(\text{not spam})}$$

   (b) Most spam filters take a naive approach and set

$$\pi(\text{spam}) = \pi(\text{not spam}) = \frac{1}{2}.$$

   If an post is known to be spam, there's an 80% chance it contains the word bitcoin. If an post is not spam, then there's a 1% chance it contains the word bitcoin. Calculate the probability the post is spam given it contains bitcoin.



HEALTH

Loyalty card data spots ovarian cancer eight months before diagnosis

Kat Lay, Health Editor

Thursday January 26 2023, 4.00pm, The Times

Fiona Murphy, a patient representative on the study, used Gaviscon for 18 months before being diagnosed with ovarian cancer

Data from shoppers' loyalty cards can be used to spot cases of ovarian cancer up to eight months early, a new study suggests.

Research involving almost 300 women revealed telltale purchasing patterns of pain and indigestion medicine among those who went on to be diagnosed with

Figure 1: An extract of the article.

(c) Suppose you take a much more pessimistic view, and assume that 80% of all posts are spam. Recalculate the probability the post is spam given it contains bitcoin.

5. A headline in The Times on 27th January 2023 read "Loyalty card data spots ovarian cancer eight months before diagnosis".

   Further on, the article says: `[The scientist] said that shopping data showed that women diagnosed with ovarian cancer had begun increasing their purchases of pain and indigestion medications up to eight months earlier. ''This suggests that long before women have recognised their symptoms as alarming enough to go to the GP, they may be treating them at home''.`

   In crude probability terms, we can think of the headline as saying we can detect cancer by investigating $\pi(\text{cancer} \mid \text{purchasing medication})$. What is the scientists saying and does that agree with the headline?