Bayesian Inference: Tutorial 1 (Week 2)

1. Drivers in age range 18-21 can be classified into 4 categories for car insurance purposes, given by:

Category	1	2	3	4
% of population in category	20%	40%	25%	15%
$\mathbb{P}$(no accidents in given year)	0.8	0.6	0.4	0.2

   Find the probability that a person chosen at random from the population has no accident in the given year. What is the probability a randomly chosen driver came from each category, given they had an accident in a given year?

2. A motorist travels regularly from St. Andrews to Edinburgh. On each occasion they choose a route at random from four possible routes. From experience, the probabilities of completing the journey in under 1.5 hours via these routes labelled 1 to 4 are 0.2, 0.5, 0.8 and 0.9, respectively. Given that on a certain occasion they complete the journey in under 1.5 hours, calculate the probability that they travelled on each of the possible routes.

3. Consider two boxes D1 and D2. The first contains 2 white and 8 red balls. The second 6 white and 4 red balls. You toss a coin and randomly pick a ball from D1 if the result is 'tails', otherwise you randomly pick a ball from D2. Let A be the event that a white ball was selected. Assuming that you forgot which box you used to pick the white ball, and that you cannot inspect the content of the boxes, find the probability that D1 was used. Similarly for D2.

4. A box of sweets contains 4 pieces of chocolate. The type of chocolate (white, milk, dark etc.) is not known, or if the pieces are of the same type of chocolate. With D_i, $i = 0, 1, 2, 3, 4$ we denote the event that the box contains i pieces of dark chocolate, and we assume, before anything is observed, that all events D_i are equally probable. We remove two pieces of chocolate and observe that they are both dark chocolate. What is the probability that there is one more piece of dark chocolate left in the box?
   (Hint: Let A denote the event that two dark pieces of chocolate are observed. Then calculate $P(A|D_i)$.)

5. A test for Hepatitis C is given to a population of possible carriers. It is believed that only 10% are positive. The test itself is only 95% accurate for people who have Hepatitis C, and 80% accurate for those who do not have the disease. What is the probability that a person has Hepatitis C given a positive result? Use this probability to derive in a sequential manner the probability that a person has Hepatitis C given two positive results. Compare your outcome with the direct calculation for the same probability for the Example in Section 1.2.1 in the lecture notes.

6. A factory has two machines for producing gluten-free vegan haggis. The old machine produces 60% and the new machine 40% of the total production.
   (a) The new machine is twice as probable as the first machine to prevent hints of wheat accidentally being added to the product. Given that a haggis does not accidentally contain hints of wheat, what is the probability that it has been produced by the new machine?
   (b) Assume a change in production so that 20% is manufactured with the old machine and 80% with the new machine. By what factor will the overall probability of the factory producing haggis without hints of wheat increase?

Bayesian Inference

Tutorial 1: Solutions

1. Let A be the event that a person at random from the population has no accident in a given year, and let B_i denote the event that an individual is from category i, for $i = 1, \ldots, 4$. Then, by definition,

$$\begin{aligned} \mathbb{P}(A) &= \sum_{i=1}^{4} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \\ &= 0.53. \end{aligned}$$

The probability that a person came from category i, given that they had an accident is given by $\mathbb{P}(B_i|A^c)$. By Bayes' Theorem,

$$\mathbb{P}(B_i|A^c) = \frac{\mathbb{P}(A^c|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A^c)}.$$

We have that, $\mathbb{P}(A^c) = 0.47$, and $\mathbb{P}(A^c|B_i) = 1 - \mathbb{P}(A|B_i)$. So that, we obtain:

Category	1	2	3	4
$\mathbb{P}(B_i\|A^c)$	0.085	0.340	0.319	0.255

2. Let A denote the event that the motorist completes the journey on time, and let B_i denote that they took the ith route. Then, we wish to calculate $\mathbb{P}(B_i|A)$. By Bayes' Theorem, we have,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

Now, we have that $\mathbb{P}(B_i) = 0.25$ for each $i = 1, \ldots, 4$, and,

$$\mathbb{P}(A) = \sum_{i=1}^{4} \mathbb{P}(A|B_i)\mathbb{P}(B_i) = 0.6.$$

Then, substituting the values into Bayes' Theorem, we obtain the probabilities,

Route	1	2	3	4
$\mathbb{P}(B_i\|A)$	0.083	0.208	0.333	0.375

3. $P(D1) = P(D2) = 0.5$. Also,

$$P(D1|A) = \frac{P(A|D1)P(D1)}{P(A|D1)P(D1) + P(A|D2)P(D2)} = \frac{(2/10)(1/2)}{(2/10)(1/2) + (6/10)(1/2)} = 1/4.$$

Thus, $P(D2|A) = 3/4$.

4. Let A denote the event that two dark pieces of chocolate are observed. Also, it is given that $P(D_i) = 1/5$ for $i = 0, 1, 2, 3, 4$. Now,

$$P(A|D_0) = 0, P(A|D_1) = 0,$$
$$P(A|D_2) = (2/4)(1/3) = 1/6,$$
$$P(A|D_3) = (3/4)(2/3) = 1/2,$$
$$P(A|D_4) = (4/4)(3/3) = 1.$$

Then,

$$P(D_3|A) = \frac{P(A|D_3)P(D_3)}{P(A)} = \frac{P(A|D_3)P(D_3)}{\sum_{i=0}^{4} P(A|D_i)P(D_i)}$$
$$= \frac{(1/2)(1/5)}{(1/5)[0 + 0 + 1/6 + 1/2 + 1]} = 3/10$$

5. Let the event that an individual has Hepatitis C be denoted by C, the event that their test result is positive be denoted by Po, and their complements be denoted C^c and Po^c respectively. Then, we have that,

$$\begin{aligned}
\mathbb{P}(C) &= 0.1 \\
\mathbb{P}(Po|C) &= 0.95 \\
\mathbb{P}(Po^c|C^c) &= 0.8 \\
\mathbb{P}(C^c) &= 0.9 \\
\mathbb{P}(Po^c|C) &= 0.05 \\
\mathbb{P}(Po|C^c) &= 0.2
\end{aligned}$$

For a true positive result,

$$\begin{aligned}
\mathbb{P}(\text{Hep C given positive test}) &= \mathbb{P}(C|Po) \\
&= \frac{\mathbb{P}(Po|C)\mathbb{P}(C)}{\mathbb{P}(Po|C)\mathbb{P}(C) + \mathbb{P}(Po|C^c)\mathbb{P}(C^c)} \\
&= \frac{0.95 \times 0.1}{(0.95 \times 0.1) + (0.2 \times 0.9)} \\
&= 0.3454545
\end{aligned}$$

Now, let $Po2$ be the event that the second test is positive. It is justifiable to assume that the two tests are independent, given the disease status and the problem specifications. The probability of disease given 2 positive tests is,

$$\mathbb{P}(C|Po, Po2) = \frac{\mathbb{P}(Po, Po2|C)\mathbb{P}(C)}{\mathbb{P}(Po, Po2|C)\mathbb{P}(C) + \mathbb{P}(Po, Po2|C^c)\mathbb{P}(C^c)}$$
$$= \frac{\mathbb{P}(Po2|Po, C)\mathbb{P}(Po|C)\mathbb{P}(C)}{\mathbb{P}(Po2|Po, C)\mathbb{P}(Po|C)\mathbb{P}(C) + \mathbb{P}(Po2|Po, C^c)\mathbb{P}(Po|C^c)\mathbb{P}(C^c)}$$
$$= \frac{\mathbb{P}(Po2|Po, C)\mathbb{P}(C|Po)\mathbb{P}(Po)\mathbb{P}(C)^{-1}\mathbb{P}(C)}{\mathbb{P}(Po2|Po, C)\mathbb{P}(C|Po)\mathbb{P}(Po)\mathbb{P}(C)^{-1}\mathbb{P}(C) + \mathbb{P}(Po2|Po, C^c)\mathbb{P}(C^c|Po)\mathbb{P}(Po)\mathbb{P}(C^c)^{-1}\mathbb{P}(C^c)}$$
$$= \frac{\mathbb{P}(Po2|C)\mathbb{P}(C|Po)}{\mathbb{P}(Po2|C)\mathbb{P}(C|Po) + \mathbb{P}(Po2|C^c)\mathbb{P}(C^c|Po)}$$
$$= \frac{0.95 \times 0.3454545}{0.95 \times 0.3454545 + 0.2 \times (1 - 0.3454545)} = 0.7148515$$

This is exactly the same probability calculated directly in the Example in the lecture notes in Section 1.2.1.

6. (a) Let A be the event that the gluten-free haggis is produced by the old machine; let A' be the event that the gluten-free haggis is produced by the new machine; let C be the event that gluten-free haggis does not accidentally contain wheat. Let $P(C|A) = k$; then $P(C|A') = 2k$, $k > 0$.

$$\begin{aligned} P(A'|C) &= \frac{P(C|A')P(A')}{P(C|A)P(A) + P(C|A')P(A')} \\ &= \frac{2k \times 0.4}{k \times 0.6 + 2k \times 0.4} = \frac{2 \times 0.4}{0.6 + 2 \times 0.4} = 4/7 \end{aligned}$$

(b) Let A be the event that the gluten-free haggis is produced by the old machine; let A' be the event that the gluten-free haggis is produced by the new machine; let C be the event that the gluten-free haggis does not accidentally contain wheat. $P(C|A) = k$ and $P(C|A') = 2k$, $k > 0$. With the current production,

$$\begin{aligned} p_1 = P(C) &= P(C|A)P(A) + P(C|A')P(A') \\ &= P(C|A)P(A) + P(C|A')P(A'), \\ &= k \times 0.6 + 2k \times 0.4 \\ &= (0.6 + 0.8)k = 1.4k \end{aligned}$$

With the new production

$$\begin{aligned} p_2 &= k \times 0.2 + 2k \times 0.8 \\ &= (0.2 + 1.6)k = 1.8k \end{aligned}$$

and $p_2/p_1 = 1.8/1.4 = 1.286$.