Module 12 self-assessment

Question 1

Suppose each of three persons flips a coin. If the outcome of one of the tosses differs from the others, the game ends, otherwise they flip their coins again. The process repeats until at least one of the coins lands on a different outcome compared to the other two. Assuming fair coins, what is the probability that the game will end after the first round of tosses? If all three coins are biased and have probability 1/4 of landing H, what is the probability that the game will end at the first round?

Solution:

The experiment involves 3 coins with 2 potential outcomes: H (heads) or T (tails), hence the sample Ω has $2^3 = 8$ random events:

$$\Omega =: \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

Two of these events are HHH and TTT where the game progresses to the next round / tossing. If the coins are fair then $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$, so as the three tosses are mutually independent the probability of each outcome in Ω above is, by the product rule, equal to $(1/2)^3 = 1/8$. By counting we then deduce

$$\mathbb{P}(\{HHH\} \cup \{TTT\}) = \mathbb{P}(\{HHH\}) + \mathbb{P}(\{TTT\}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4},$$

where we have used that events TTT and HHH are disjoint. Hence the probability that the game ends after the first round is the probability of the complement of the $HHH \cup TTT$ which is

$$\mathbb{P}\big((HHH \cup TTT)^c\big) = 1 - \frac{1}{4} = \frac{3}{4}.$$

If the coins are biased with $\mathbb{P}(H) = 1/4$, and hence $\mathbb{P}(T) = 3/4$ then we can adjust the calculations above

$$\mathbb{P}(HHH) = (1/4)^3 = 1/64, \quad \mathbb{P}(TTT) = (3/4)^3 = 27/64.$$

From this the probability of having the same outcome in all three tosses becomes

$$\mathbb{P}(HHH \cup TTT) = \mathbb{P}(HHH) + \mathbb{P}(TTT) = 28/64$$

and thus the probability of ending the game from the first round with the biased coins is $1 - 28/64 = 36/64$, which is lower than the unbiased case above.

Question 2

It is estimated that 50% of received emails are spam emails. A software has been installed by the IT department of the University to filter these spam emails before they reach our inboxes. The supplying software company claims that it can detect 99% of spam emails, and it is believed that the probability for a false positive (a non-spam email classified as spam) is only 5%. If an email is detected as spam, what is the probability that it is in fact a non-spam email? [Hint: Use Bayes' theorem.]

Solution:

To solve this we will rely on Bayes' theorem, taking into account the prior belief about the false positive probability. To begin with we define the events A = an email is detected as spam, B = an email is spam, and its complement B^c = an email is not spam. We can now express the first sentence in the question "It is estimated that 50% of received emails are spam emails" as $\mathbb{P}(B) = 0.5$ and thus it follows that $P(B^c) = 0.5$, i.e. "50% of received emails are not spam emails".

We can then translate the company's claim that "it can detect 99% of spam emails" as $\mathbb{P}(A|B) = 0.99$ which is the same as saying "it can detect 99% of spam emails *if they are spam.* Note that the event in italics (here B) comes second in the conditional probability. Then it says "the probability for a false positive is only 5%", that is the probability of being detected as spam *if they are not spam*" is $\mathbb{P}(A|B^c) = 0.05$. Now to the required: "the probability that an email is non-spam *if it is detected as spam*" which we can write as $\mathbb{P}(B^c|A)$. Since we have $\mathbb{P}(A|B^c)$ and we want $\mathbb{P}(B^c|A)$ then we can use Bayes's theorem to get

$$\begin{aligned}
\mathbb{P}(B^c|A) &= \frac{\mathbb{P}(A|B^c)\mathbb{P}(B^c)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(A|B^c)\mathbb{P}(B^c)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} \\
&= \frac{0.05 \cdot 0.5}{0.05 \cdot 0.5 + 0.99 \cdot 0.5} = \frac{5}{104}
\end{aligned}$$

and the expansion for $\mathbb{P}(A)$ at the denominator follows from the rule of total probability.