Probability and Statistics for Data Analytics – 2025/26
Problem Sheet 3

The questions on this sheet are based on the material on conditional probability from Week 3 lectures. This sheet is not for assessment. We will discuss selected questions from it in the Week 4 seminar.

---

1. I have three coins in my pocket. Two of them are ordinary fair coins, the third has Heads on both sides.

   (a) I take a coin at random from my pocket and toss it. What is the probability that it comes up Heads?

   (b) Given that it did come up Heads, what is the conditional probability that it is the double-headed coin?

   (c) Given that it did come up Heads, what is the conditional probability that a second toss of the same coin also comes up Heads?

   (d) Given that two tosses of the same coin both come up Heads, what is the conditional probability that it is the double-headed coin? [Before doing this think do you expect the answer to be larger or smaller than the answer to part (b)? Why?]

   (e) Suppose that 100 tosses of the same coin show a Head every time. Without doing any more calculations, say roughly what you expect the conditional probability that it is the double-headed coin to be? How would you explain this in non-mathematical terms?

2. A certain medical condition affects 1% of a population. A new AI tool for detecting this condition from a scan has been developed. It has a 95% success rate at correctly detecting that a person with the condition has it, and only a 2% chance of incorrectly deciding that a healthy person has the condition. A randomly chosen person from the population undertakes this test and the test shows positive.

   (a) What is the probability that they do have the condition?

   (b) A politician suggests that this test could be used for a national screening programme. What insight into the possible disadvantages of doing this does the calculation of part (a) provide? How could these disadvantages be mitigated?

3. Suppose that A and B are events with $P(A) > 0$ and $\mathbb{P}(B) > 0$ and that $\mathbb{P}(A \mid B) > \mathbb{P}(A)$.

(a) What can you say about $\mathbb{P}(B \mid A)$?

(b) How would you explain the relationship between events A and B which satisfy this property in non-mathematical language?

4.

(a) Suppose that T is a discrete random variable measuring the number of days until some event happens. We say that T has a Geometric(p) distribution if it takes values in $1, 2, 3, \ldots$ and the pmf is $\mathbb{P}(T = k) = (1-p)^{k-1}p$.

(i) What is $\mathbb{P}(T > k)$?

(ii) Show that for any $a, k \in \mathbb{N}$ we have $\mathbb{P}(T > a + k \mid T > a) = \mathbb{P}(T > k)$.

(b) Suppose that the lifetime of a component is a continuous random variable V which follows an Exponential distribution with parameter λ.

(i) What is $\mathbb{P}(V > k)$?

(ii) Show that for any $a, t \in \mathbb{R}$ we have $\mathbb{P}(V > a + k \mid V > a) = \mathbb{P}(V > k)$.

(c) Why do you think the results in parts (a)(ii) and (b)(ii) are sometimes called the *memoryless property* of the Geometric and Exponential distributions.

5. Read this article from mathematical epidemiologist Adam Kucharski's blog:

`https://kucharski.substack.com/p/small-hallucinations-big-problems`

(a) Using what you have learnt in this module, redo this analysis in the case that the probability of the unusual event we are looking for is p (rather than the 1 in 1000 that the article uses).

(b) Do you think the author did a good job at explaining the mathematics to a non-mathematical audience?

(c) Look at the Student Forum on the module QMplus page.

`https://qmplus.qmul.ac.uk/mod/forum/view.php?id=3237993`

Make a comment (which could be about your answer to parts (a) or (b) of this question or something else) on the discussion thread about this article.

Robert Johnson
`r.johnson@qmul.ac.uk`