This homework runs from Thursday 7 November 2024 until 12 noon on Thursday 14 November 2024. Submission is to Gradescope Homework 5.

You should aim to write out solutions that someone who does not already know the answer could follow and understand.

Please remember the good scholarly practice requirements of the University regarding work for credit. You can find guidance at the School page https://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct. This also has links to the relevant University pages.

## Question 1

Suppose that a pharmaceutical company has developed a new non-invasive test for a type of cancer. Their studies show that this new test has the following properties.

- If the test is performed on a (random) person who has this type of cancer, then there is an 88% chance that the test result will be positive.
- If, on the other hand, the test is performed on a (random) person who does not have this cancer, then there is a 9% chance that the test result will be positive.

Approximately 1 in 1000 persons in the entire population have this type of cancer.

Suppose that this new test is performed on a (random) person in the population. What is the probability that the person actually has this cancer, given that their test result was positive?

[*3 marks*]

*This ends the part of the question to which you have to submit a written answer. The following questions are for you to think through now and then discuss at the tutorial*

Considering the probability you have just calculated, suppose you were advising NHS Scotland regarding whether or not to recommend this test for the general population. Would it be worthwhile? What would you say to them?

Suppose also that the test is available at a scale which could cover the whole population of Scotland, but at a cost of £350 for each test. Does this affect your advice?

Suppose further that if this test is positive it requires additional invasive tests to see if the cancer is indeed present; and these follow-up tests are actually quite painful and cause prolonged discomfort for patients, whether or not they actually have the cancer.

What would your recommendation now be to NHS Scotland regarding whether to regularly carry out this test, and why? There is no right answer to this, but it gives you an idea of trade-offs routinely arising in population health care.

**Turn to the next page for Question 2**

**Question 2**

Suppose a particular web server provides generally reliable service but does occasionally fail. The time between failures is a random variable $T$ measured in days with exponential distribution $T \sim \text{Exp}(0.01)$. This means it has the following probability density function (PDF) and cumulative distributions function (CDF), where $\lambda = 0.01$.

$$\text{PDF } f(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \qquad \text{CDF } F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

**(a)** What is $\text{E}(T)$, the *mean time between failures* (MTBF) in days?

**(b)** Calculate the chance that the server runs for at least two weeks without failure.

**(c)** Calculate $P(14 < T \,\&\, T < 21)$, the probability that the server first fails in the third week of operation.

**(d)** Show how to use the previous results to work out the conditional probability $P(T < 21 \mid T > 14)$.

**(e)** Use the previous results and the distinctive property of the exponential distribution to calculate the conditional probability that the server fails in the fifth week of operation, given that it has already worked for the first four weeks.

For each part explain your answer and show your working. [*7 marks*]

**Solution 1**

We want to compute $P(C|T)$, where $C$ represents the event that the person has cancer, and $T$ represents the event that the test is positive. We use $\overline{C}$ to denote the event that the person does not have cancer.

Bayes' theorem tells us that $P(C|T) = P(T|C)P(C)/P(T)$, and the Law of Total Probability gives us the following.

$$\begin{aligned} P(T) &= P(T \cap C) + P(T \cap \overline{C}) \\ &= P(T|C)P(C) + P(T|\overline{C})P(\overline{C}) \end{aligned}$$

Therefore:

$$\begin{aligned} P(C|T) &= \frac{P(T|C)P(C)}{P(T)} \\ &= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\overline{C})P(\overline{C})} \\ &= \frac{(0.88)(0.001)}{(0.88)(0.001) + (0.09)(0.999)} = 0.00969 \quad (3\,\text{s.f.}) \end{aligned}$$

Thus, a random person who tests positive for this test is still very unlikely to have this type of cancer: they have a chance of less than 1 in 100 of having this cancer.

Still, this nearly 10 times more likely than the probability (1/1000) that a random person in the general population has this cancer. So the test does tell us something.

However, if the follow-up tests are painful and cause substantial long term discomfort, it is tricky to justify causing so much discomfort to so many people in the population who do not have the cancer, in order to catch early on the rarer cases of those who do have the cancer. Furthermore, obviously in the real world, the costs of such tests also have to be taken into account.

Although this example is very simplistic, it already indicates some of the dilemmas and trade-offs faced regularly by policy-makers in health care.

**Solution 2**

**(a)** $\mathrm{E}(T) = 1/\lambda = 100$. This is standard for the exponential distribution, but also reasonable to calculate from scratch using integration by parts.

$$\begin{aligned} \mathrm{E}(T) &= \int_{-\infty}^{+\infty} t \cdot f(t)\, dt = \int_{0}^{+\infty} \lambda t e^{-\lambda t}\, dt \\ &= \left[-te^{-\lambda t}\right]_0^\infty - \int_0^\infty -e^{-\lambda t}\, dt = 0 - \left[\frac{1}{\lambda}e^{-\lambda t}\right]_0^\infty = \frac{1}{\lambda} = 100 \end{aligned}$$

**(b)** $P(T > 14) = 1 - F(14) = e^{-0.14} = 0.87 \ (2\,\text{s.f.})$

**(c)** $P(14 < T \,\&\, T < 21) = F(21) - F(14) = e^{-0.14} - e^{-0.21} = 0.059 \ (2\,\text{s.f.})$

**(d)**

$$P(T < 21 \mid T > 14) = \frac{P(T < 21 \,\&\, T > 14)}{P(T > 14)} = \frac{e^{-0.14} - e^{-0.21}}{e^{-0.14}} = 0.068 \quad (2\,\text{s.f.})$$

**(e)** By the memoryless property of the exponential distribution this probability is the same as the previous one, 0.068.

*Note:* The question here specifically asks for a solution using the distinctive property of the exponential distribution (i.e. that it is memoryless) and the previous results (in this case, the answer to part (d), which is exactly the value required). Alternate routes that involve further calculation might get the right numerical value but are not solutions to the problem as stated.