Probability and Statistics for Data Analytics – 2025/26
Problem Sheet 2

The questions on this sheet are based on the material on continuous random variables from Week 2 lectures. This sheet is not for assessment. We will discuss selected questions from it in the Week 3 seminar.

1. A continuous random variable X has pdf of the form

$$f_X(x) = \begin{cases} 0 & \text{if } x < \frac{1}{2} \\ \frac{C}{x^2} & \text{if } x \geqslant \frac{1}{2}. \end{cases}$$

for some constant C.

(a) What is that value of C?

(b) What is $f_X(\frac{1}{2})$?

(c) Find the cdf of X?

(d) Write down a couple of ways you could check your answer to (c) is plausible. Perform those checks and revisit your answer to (c) if necessary.

(e) How would you calculate $\mathbb{P}(1 < X < 3)$ using the cdf?

(f) How would you calculate $\mathbb{P}(1 < X < 3)$ using the pdf?

2. A continuous random variable X has cdf

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{6} & \text{if } 0 \leqslant x < 3 \\ \frac{1}{2} & \text{if } 3 \leqslant x < 4 \\ \frac{x-2}{4} & \text{if } 4 \leqslant x < 6 \\ 1 & \text{if } x \geqslant 6. \end{cases}$$

(a) Find the pdf of X.

(b) Write down expressions for $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ and compute them.

(c) Describe the distribution this random variable follows in words in a way that would make sense to a non-mathematician.

3. The number $m \in \mathbb{R}$ is a *median* for the random variable X if $F_X(m) = \frac{1}{2}$.

(a) Indicate why every continuous random variable has a median. [This can be an informal explanation rather than a rigorous mathematical proof, although if you know some analysis you could try to write a proof.]

(b) Can you say anything about $f_X(m)$ (i.e. the pdf evaluated at m)?

4. Let T be the random variable giving the time (in minutes) between consecutive customers ariving in a shop. Suppose that $T \sim \text{Exp}(0.5)$. Each customer spends 5 minutes in the shop and then leaves. The first customer of the day has just entered the shop.

(a) What is the probability that the next customer does not arrive until after the first customer has left.

(b) What is the expectation of the time before the next customer arrives.

(c) Is the median of the time before the next customer arrives smaller, larger or the same as the expectation.

(d) What is the probability that the time before the next customer arrives is greater than twice its expectation.

(e) How do your answers to (c) and (d) change if the parameter of T changes?

5. Let Ω be the right-angled triangle with vertices at $(0,0), (1,0)$ and $(1,1)$. Let a be a point chosen randomly from within Ω with the probability that a is in any fixed region being proportional to the area of the region. Let T be the random variable "The angle that the line from $(0,0)$ to a makes with the x-axis. Find the cdf and pdf of T?

6. Decide whether each of the following statements is true or false. Give a reason in either case:

(a) For any continuous random variable X, the values of the pdf are probabilities so $0 \leqslant f_X(x) \leqslant 1$ for all $x \in \mathbb{R}$.

(b) For any continuous random variable, the median is always unique (i.e. there is only one $m \in \mathbb{R}$ which satisfies the definition of median).

Robert Johnson
r.johnson@qmul.ac.uk