STAT1301 Advanced Analysis of Scientific Data

Semester 2, 2025

Assignment 2

If X is a random variable of the sample space Ω, an abbreviation of set notation is as follows:

$$\text{Abbreviate } \{d : \forall d \in \Omega \text{ and } \mathrm{X}(d) = x\} \tag{1}$$
$$\text{As } \{\mathrm{X} = x\} \tag{2}$$

Additionally, when thinking in terms of sets becomes obselete,

$$\text{Abbreviate } \mathrm{P}(\{d : \forall d \in \Omega \text{ and } \mathrm{X}(d) = x\}) \tag{3}$$
$$\text{As } \mathrm{P}(\mathrm{X} = x) \tag{4}$$

This abbriviation will be used with inequalities as well

1. Question 1

   To begin, lets define the sample space

   $$\Omega = \{(a, b) \in \{1, 2, 3, 4, 5, 6\}\}$$
   $$|\Omega| = 36$$

   Notice this is uniform, and hence that a and b are independant

   $$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{36} \tag{5}$$

   a)

   Let X be a random variable representing the payout of a given dice roll $(a, b) \in \Omega$:

   $$\mathrm{X}((a, b) \in \Omega) = a \cdot b$$

   Let f_{X} be the PMF of X. By cases, the probability distribution of X can be deduced:

   $$f_{\mathrm{X}}(1) = \frac{1}{36} \qquad f_{\mathrm{X}}(8) = \frac{2}{36} \qquad f_{\mathrm{X}}(18) = \frac{2}{36}$$
   $$f_{\mathrm{X}}(2) = \frac{2}{36} \qquad f_{\mathrm{X}}(9) = \frac{1}{36} \qquad f_{\mathrm{X}}(20) = \frac{2}{36}$$
   $$f_{\mathrm{X}}(3) = \frac{2}{36} \qquad f_{\mathrm{X}}(10) = \frac{2}{36} \qquad f_{\mathrm{X}}(24) = \frac{2}{36}$$
   $$f_{\mathrm{X}}(4) = \frac{3}{36} \qquad f_{\mathrm{X}}(12) = \frac{4}{36} \qquad f_{\mathrm{X}}(25) = \frac{1}{36}$$
   $$f_{\mathrm{X}}(5) = \frac{2}{36} \qquad f_{\mathrm{X}}(15) = \frac{2}{36} \qquad f_{\mathrm{X}}(30) = \frac{2}{36}$$
   $$f_{\mathrm{X}}(6) = \frac{4}{36} \qquad f_{\mathrm{X}}(16) = \frac{1}{36} \qquad f_{\mathrm{X}}(36) = \frac{1}{36}$$

For all other values x, $f_{\mathrm{X}}(x) = 0$

b)

$$\begin{aligned}
\mathrm{E}(\mathrm{X}) &= \sum_{d\in\Omega} \mathrm{X}(d)\mathrm{P}(d) && (6)\\
&= \sum_{x\in\mathrm{Domain}[\mathrm{X}]} x\mathrm{P}(\{\mathrm{X}=x\}) && (7)\\
&= 1\cdot f_{\mathrm{X}}(1) + 2\cdot f_{\mathrm{X}}(2) + \cdots 30\cdot f_{\mathrm{X}}(30) + 36\cdot f_{\mathrm{X}}(36) && (8)\\
&= \frac{1}{36} + \frac{4}{36} + \cdots \frac{60}{36} + \frac{36}{36} && (9)\\
&= \frac{441}{36} = \frac{49}{4} = 12.25 && (10)
\end{aligned}$$

2. A laboratory is studying the growth of algae in a controlled environment. The growth of algae is measured by the amount of biomass produced (in grams), which can be modelled by a random variable X with probability density function (pdf)

$$f_X(x) = c\left(x^2 - 60x + 800\right) \quad \text{for } 0 < x < 20,$$

where c is a constant.

   (a) Find the value of c. [2 marks]

   (b) Find the cumulative distribution function (cdf) of X. [3 marks]

   (c) Find the expected value of (X). [2 marks]

   (d) The laboratory equipment can only detect biomass that exceeds 2 grams (the minimum detectable amount). What is the probability that the biomass exceeds 10 grams, given that it is detectable? [3 marks]

3. Each day, a quality control officer inspects a random sample of 25 products from the production line of a factory. The probability of a product passing inspection (being defect-free) is 0.25. If the product passes, the factory saves \$3 in repair costs. If the product fails, the factory incurs an additional \$1 cost for re-inspection after repair. Let X be the number of products that passes inspection on a given day, and Y be the net savings for the factory on that day.

$$\sigma$$

   (a) State the distribution of X, including all its parameters. [2 marks]

   (b) What is the minimum sample size needed so that the probability of finding at least one defect-free product exceeds 99%? [3 marks]

   (c) Calculate the expected value and variance of the factory's net savings. [3 marks]

   (d) What is the probability that the factory will save at least \$27 on a given day? [2 marks]

4. A storeroom in a warehouse maintains strict temperature control to ensure that sensitive materials are stored at optimal conditions. The temperature of the storeroom follows a normal distribution with mean μ and standard deviation σ degrees Celsius (°C). The storeroom has a temperature threshold of 8°C to avoid damaging the materials.

   You may use statistical tables to answer this question, then use R to verify your results.

   (a) Suppose the storeroom temperature is adjusted so that $\mu = 7.5$°C and $\sigma = 0.3$°C. What is the probability that the temperature of the storeroom will be between 7.2°C and 8°C?

   [3 marks]

   (b) Assume that $\sigma = 0.3$°C. What should μ be set to so that the storeroom temperature exceeds 8°C only 1% of the time? [3 marks]

   (c) What is the largest standard deviation σ that will keep the temperature within 1°C of the mean with 95% probability? [4 marks]