STAT1301 Advanced Analysis of Scientific Data
Semester 2, 2025
Assignment 2

1 Introduction

Throughout the report, the following syntactical shortcuts and notation will be used.

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,

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

The abbriviation will be used with inequalities as well.

Given some random variable X, there must exist a function mapping from the sample space Ω to the domain of X, which can be at most $\mathbb{R}$. This function is (intuitively) named X. This function incidentally defines the random variable, which is the motivating reason for using its letter to represent its mapping. The notation Domain[X] will be used throughout this report to indicate the domain of the function mapping X and hence the random variable X itself by definition.

2 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

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

2.1 Part 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. Note $f_{\mathrm{X}}(x \in \Omega) = \mathrm{P}(\{\mathrm{X} = x\})$. By cases, the probability distribution of X can be deduced:

$$f_{\mathrm{X}}(1) = \frac{1}{36} \quad f_{\mathrm{X}}(8) = \frac{2}{36} \quad f_{\mathrm{X}}(18) = \frac{2}{36}$$

$$f_{\mathrm{X}}(2) = \frac{2}{36} \quad f_{\mathrm{X}}(9) = \frac{1}{36} \quad f_{\mathrm{X}}(20) = \frac{2}{36}$$

$$f_{\mathrm{X}}(3) = \frac{2}{36} \quad f_{\mathrm{X}}(10) = \frac{2}{36} \quad f_{\mathrm{X}}(24) = \frac{2}{36}$$

$$f_{\mathrm{X}}(4) = \frac{3}{36} \quad f_{\mathrm{X}}(12) = \frac{4}{36} \quad f_{\mathrm{X}}(25) = \frac{1}{36}$$

$$f_{\mathrm{X}}(5) = \frac{2}{36} \quad f_{\mathrm{X}}(15) = \frac{2}{36} \quad f_{\mathrm{X}}(30) = \frac{2}{36}$$

$$f_{\mathrm{X}}(6) = \frac{4}{36} \quad f_{\mathrm{X}}(16) = \frac{1}{36} \quad f_{\mathrm{X}}(36) = \frac{1}{36}$$

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

2.2 Part b)

This makes determining the expected value of X trivial:

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

2.3 Part c)

Evaluating Var(X) is similarly trivial

$$\begin{aligned}
\mathrm{Var}(\mathrm{X}) &= \mathrm{E}[(\mathrm{X}-\mathrm{E}(X))^2] && (11)\\
&= \sum_{c\in\Omega} (\mathrm{X}(c) - \frac{49}{4})^2\mathrm{P}(\{c\}) && (12)\\
&= \sum_{x\in\mathrm{Domain[X]}} (x-\frac{49}{4})^2\mathrm{P}(\{\mathrm{X}=x\}) && (13)\\
&= (1-\frac{49}{4})^2\cdot\frac{1}{36} + (2-\frac{49}{4})^2\cdot\frac{2}{36} + \cdots (30-\frac{49}{4})^2\cdot\frac{2}{36} + (36-\frac{49}{4})^2\cdot\frac{1}{36} && (14)\\
&= \frac{11515}{144} \approx 79.97 && (15)\\
\implies \sigma_{\mathrm{X}} &= \sqrt{\mathrm{Var}(\mathrm{X})} = \sqrt{\frac{11515}{144}} \approx 8.942 && (16)
\end{aligned}$$

3 Question 2

Understanding this question in terms of a sample space isn't very fruitful. Ω is completely unspecified, we can only deduce that $|\Omega| \geq (0, 20)$, which implies it is continuous. $\mathrm{P}(A) : \exists A \in \Omega$ is also completely unknown.

3.1 Part a)

Let X be the continuous random variable of algae growth as measured in grams of biomass produced. Note Domain[X] $= (0, 20)$.

Since X is a random variable, its PDF f_{X} must sum to 1:

$$
\begin{aligned}
1 &= \int_{c\in\Omega} \mathrm{P}(\{c\}) && (17)\\
&= \int_{x\in\mathrm{Domain}[X]} \mathrm{P}(\{\mathrm{X} = x\}) && (18)\\
&= \int_0^{20} c(x^2 - 60x + 800)\mathrm{d}x && (19)\\
&= c[\frac{1}{3}x^3 - 30x^2 + 800x]_{x=0}^{x=20} && (20)\\
1/c &= [\frac{1}{3}(20)^3 - 30(20)^2 + 800(20)] - [0 - 0 + 0] && (21)\\
1/c &= \frac{20000}{3} && (22)\\
c &= \frac{3}{20000} && (23)
\end{aligned}
$$

3.2 Part b)

Let F_{X} be the CDF of X:

$$
\begin{aligned}
F_{\mathrm{X}} &= \int_{-\infty}^{x} f_{\mathrm{X}}(x)\mathrm{d}x && (24)\\
&= \int_0^{x} c(x^2 - 60x + 800)\mathrm{d}x && (25)\\
&= c[\frac{1}{3}x^3 - 30x^2 + 800x]_{x=0}^{x=x} && (26)\\
\frac{F_{\mathrm{X}}}{c} &= [\frac{1}{3}x^3 - 30x^2 + 800x] - [\frac{1}{3}0^3 - 30\cdot 0^2 + 800\cdot 0] && (27)\\
\implies F_{\mathrm{X}} &= c(\frac{1}{3}x^3 - 30x^2 + 800x) \text{ for } 0 \leq x \leq 20 && (28)\\
&= \frac{1}{20000}x^3 - \frac{9}{2000}x^2 + \frac{3}{25}x && (29)
\end{aligned}
$$

3.3 Part c)

$$
\begin{aligned}
\mathrm{E(X)} &= \int_{x\in\mathrm{Domain[X]}} x f_{\mathrm{X}} \mathrm{d}x && (30)\\
&= \int_0^{20} x \cdot c(x^2 - 60x + 800)\mathrm{d}x && (31)\\
\frac{\mathrm{E(X)}}{c} &= \int_0^{20} x^3 - 60x^2 + 800x\mathrm{d}x && (32)\\
&= [\frac{1}{4}x^4 - 20x^3 + 400x^2]_{x=0}^{x=20} && (33)\\
&= [\frac{1}{4}(20)^4 - 20(20)^3 + 400(20)^2] - [0 - 0 + 0] && (34)\\
&= 40000 - 160000 + 160000 && (35)\\
\mathrm{E(X)} &= c \cdot 40000 && (36)\\
\mathrm{E(X)} &= 6 \text{ grams} && (37)
\end{aligned}
$$

3.4 Part d)

$$
\begin{aligned}
&\mathrm{P}(\{\mathrm{X} > 10\}|\{\mathrm{X} > 2\}) && (38)\\
&= \frac{\mathrm{P}(\{\mathrm{X} > 10\} \cap \{\mathrm{X} > 2\})}{\{\mathrm{X} > 2\}} && (39)\\
&= \frac{\mathrm{P}(\{\mathrm{X} > 10\})}{\mathrm{P}(\{\mathrm{X} > 2\})} && (40)
\end{aligned}
$$

From the CDF definition of X, $\mathrm{P}(\{\mathrm{X} < x\}) = f_{\mathrm{X}}(x)$

$$
\begin{aligned}
\implies \mathrm{P}(\{\mathrm{X} > 10\}) &= 1 - \mathrm{P}(\{\mathrm{X} < 10\}) && (41)\\
&= 1 - F_{\mathrm{X}}(10) && (42)\\
&= 1 - \frac{4}{5} && (43)\\
&= \frac{1}{5} && (44)
\end{aligned}
$$

$$
\begin{aligned}
\implies \mathrm{P}(\{\mathrm{X} > 2\}) &= 1 - \mathrm{P}(\{\mathrm{X} < 2\}) && (45)\\
&= 1 - F_{\mathrm{X}}(2) && (46)\\
&= 1 - \frac{139}{625} && (47)\\
&= \frac{486}{625} && (48)
\end{aligned}
$$

$$
\implies \frac{\mathrm{P}(\{\mathrm{X} > 10\})}{\mathrm{P}(\{\mathrm{X} > 2\})} = \frac{\frac{1}{5}}{\frac{486}{625}} = \frac{125}{486} \approx 0.2572 \qquad (49)
$$

Therefore, the probability that the biomass exceeds 10 grams, given that it is detectable, is above in (49) $= \frac{125}{486}$.