

# 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  $\Omega$ , an abbreviation of set notation is as follows:

$$\text{Abbreviate } \{d : \forall d \in \Omega \text{ and } X(d) = x\} \quad (1)$$

$$\text{As } \{X = x\} \quad (2)$$

Additionally, when thinking in terms of sets becomes obsolete,

$$\text{Abbreviate } P(\{d : \forall d \in \Omega \text{ and } X(d) = x\}) \quad (3)$$

$$\text{As } P(X = x) \quad (4)$$

The abbreviation will be used with inequalities as well.

Given some random variable  $X$ , there must exist a function mapping from the sample space  $\Omega$  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  $\text{Domain}[X]$  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, let's 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 independent

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

### 2.1 Part a)

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

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

Let  $f_X$  be the PMF of  $X$ . Note  $f_X(x \in \Omega) = P(\{X = x\})$ . By cases, the probability distribution of  $X$  can be deduced:

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

For all other values  $x$ ,  $f_X(x) = 0$

## 2.2 Part b)

This makes determining the expected value of  $X$  trivial:

$$E(X) = \sum_{c \in \Omega} X(c)P(c) \quad (6)$$

$$= \sum_{x \in \text{Domain}[X]} xP(\{X = x\}) \quad (7)$$

$$= 1 \cdot f_X(1) + 2 \cdot f_X(2) + \cdots 30 \cdot f_X(30) + 36 \cdot f_X(36) \quad (8)$$

$$= \frac{1}{36} + \frac{4}{36} + \cdots \frac{60}{36} + \frac{36}{36} \quad (9)$$

$$= \frac{441}{36} = \frac{49}{4} = 12.25 \quad (10)$$

## 2.3 Part c)

Evaluating  $\text{Var}(X)$  is similarly trivial

$$\text{Var}(X) = E[(X - E(X))^2] \quad (11)$$

$$= \sum_{c \in \Omega} (X(c) - \frac{49}{4})^2 P(\{c\}) \quad (12)$$

$$= \sum_{x \in \text{Domain}[X]} (x - \frac{49}{4})^2 P(\{X = x\}) \quad (13)$$

$$= (1 - \frac{49}{4})^2 \cdot \frac{1}{36} + (2 - \frac{49}{4})^2 \cdot \frac{2}{36} + \cdots (30 - \frac{49}{4})^2 \cdot \frac{2}{36} + (36 - \frac{49}{4})^2 \cdot \frac{1}{36} \quad (14)$$

$$= \frac{11515}{144} \approx 79.97 \quad (15)$$

$$\Rightarrow \sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\frac{11515}{144}} \approx 8.942 \quad (16)$$

### 3 Question 2

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

Since  $X$  is a random variable, its PDF  $f_X$  must sum to 1:

$$1 = \int_{c \in \Omega} P(\{c\}) \quad (17)$$

$$= \int_{x \in \text{Domain}[X]} P(\{X = x\}) \quad (18)$$

$$= \int_0^{20} c(x^2 - 60x + 800)dx \quad (19)$$

$$= c \left[ \frac{1}{3}x^3 - 30x^2 + 800x \right]_{x=0}^{x=20} \quad (20)$$

$$1/c = \left[ \frac{1}{3}(20)^3 - 30(20)^2 + 800(20) \right] - [0 - 0 + 0] \quad (21)$$

$$1/c = \frac{20000}{3} \quad (22)$$

$$c = \frac{3}{20000} \quad (23)$$

#### 3.2 Part b)

Let  $F_X$  be the CDF of  $X$ :

$$F_X = \int_{-\infty}^x f_X(x)dx \quad (24)$$

$$= \int_0^x c(x^2 - 60x + 800)dx \quad (25)$$

$$= c \left[ \frac{1}{3}x^3 - 30x^2 + 800x \right]_{x=0}^{x=x} \quad (26)$$

$$\frac{F_X}{c} = \left[ \frac{1}{3}x^3 - 30x^2 + 800x \right] - \left[ \frac{1}{3}0^3 - 30 \cdot 0^2 + 800 \cdot 0 \right] \quad (27)$$

$$\implies F_X = c \left( \frac{1}{3}x^3 - 30x^2 + 800x \right) \text{ for } 0 \leq x \leq 20 \quad (28)$$

$$= \frac{1}{20000}x^3 - \frac{9}{2000}x^2 + \frac{3}{25}x \quad (29)$$

### 3.3 Part c)

$$E(X) = \int_{x \in \text{Domain}[X]} x f_X dx \quad (30)$$

$$= \int_0^{20} x \cdot c(x^2 - 60x + 800) dx \quad (31)$$

$$\frac{E(X)}{c} = \int_0^{20} x^3 - 60x^2 + 800x dx \quad (32)$$

$$= \left[ \frac{1}{4}x^4 - 20x^3 + 400x^2 \right]_{x=0}^{x=20} \quad (33)$$

$$= \left[ \frac{1}{4}(20)^4 - 20(20)^3 + 400(20)^2 \right] - [0 - 0 + 0] \quad (34)$$

$$= 40000 - 160000 + 160000 \quad (35)$$

$$E(X) = c \cdot 40000 \quad (36)$$

$$E(X) = 6 \text{ grams} \quad (37)$$

### 3.4 Part d)

$$P(\{X > 10\} | \{X > 2\}) \quad (38)$$

$$= \frac{P(\{X > 10\} \cap \{X > 2\})}{P(\{X > 2\})} \quad (39)$$

$$= \frac{P(\{X > 10\})}{P(\{X > 2\})} \quad (40)$$

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

$$\implies P(\{X > 10\}) = 1 - P(\{X < 10\}) \quad (41)$$

$$= 1 - F_X(10) \quad (42)$$

$$= 1 - \frac{4}{5} \quad (43)$$

$$= \frac{1}{5} \quad (44)$$

$$\implies P(\{X > 2\}) = 1 - P(\{X < 2\}) \quad (45)$$

$$= 1 - F_X(2) \quad (46)$$

$$= 1 - \frac{139}{625} \quad (47)$$

$$= \frac{486}{625} \quad (48)$$

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