

### Question 1

$$a) f_x(x) = \int_0^1 f_{x,y}(x,y) dy \rightarrow$$

$$\rightarrow f_x(x) = \int_0^1 \frac{12}{49} (2+x+xy+4y^2) dy \rightarrow$$

$$\rightarrow f_x(x) = \frac{12}{49} \left[ 2y + xy + \frac{xy^2}{2} + \frac{4y^3}{3} \right]_0^1 \rightarrow$$

$$\rightarrow f_x(x) = \frac{12}{49} \left[ 2 + x + \frac{x}{2} + \frac{4}{3} \right] \rightarrow$$

$$\rightarrow f_x(x) = \frac{12}{49} \left[ \frac{10}{3} + \frac{3x}{2} \right] \rightarrow$$

$$\rightarrow f_x(x) = \frac{40}{49} + \frac{18x}{49} = \boxed{\frac{18x+40}{49}} \rightarrow \text{answer}$$

---

b)

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y) dx \rightarrow$$

$$\rightarrow f_Y(y) = \int_0^1 \frac{12}{49} (2+x+xy+4y^2) dx \rightarrow$$

$$\rightarrow f_Y(y) = \frac{12}{49} \left[ 2x + \frac{x^2}{2} + \frac{x^2 y}{2} + 4y^2 x \right]_0^1 \rightarrow$$

$$\rightarrow f_Y(y) = \frac{12}{49} \left[ 2 + \frac{1}{2} + \frac{1}{2}y + 4y^2 \right] \rightarrow$$

$$\rightarrow f_Y(y) = \frac{12}{49} \left[ \frac{5}{2} + \frac{1}{2}y + 4y^2 \right] \rightarrow$$

$$\rightarrow f_Y(y) = \frac{30}{49} + \frac{6}{49}y + \frac{48}{49}y^2 =$$

$$\boxed{\frac{48y^2 + 6y + 30}{49}}$$

↓  
answer

(c)

Two random variables,  $X$  and  $Y$ , are independent if and only if their joint probability function  $f_{X,Y}(x,y)$  can be expressed as the product of their individual probability functions,  $f_X(x)$  and  $f_Y(y)$ . In other words, for independent variables, the joint density function equals the product of the marginal density functions.

for independence:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

We are given  $f_{X,Y}(x,y) = \frac{12}{49} (2+x+xy+4y^2)$

We found  $f_X(x) = \frac{18x+40}{49}$  and

$$f_Y(y) = \frac{48y^2 + 6y + 30}{49}$$

calculation:

$$f_X(x) \cdot f_Y(y) = \frac{(18x+40)(48y^2 + 6y + 30)}{49^2} \neq f_{X,Y}(x,y)$$

Since  $f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y)$  for all  $x$  and  $y$  within their ranges,  $X$  and  $Y$  are not independent variables.

## Question 2

a)

the possible values of  $X$ :  $x=1, 2, 3, 4, 5, 6$

the possible values of  $Y$ : depend on the value of  $X$

$$f_{X,Y}(x,y) = P(X=x \& Y=y)$$

The number of heads obtained when flipping  $x$  coins follows a binomial distribution with parameters  $x$  (number of trials) and  $p = 1/2$  (prob. of success in each trial since it's a fair coin). In this case,  $Y$  follows a binomial distribution.

The joint PMF can be represented as:

$$f_{X,Y}(x,y) = \begin{cases} \left(\frac{1}{6}\right) \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x & \text{if } 0 \leq y \leq x \& 1 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

$\binom{x}{y}$  → represents the number of ways to choose  $y$  successes (heads) out of  $x$  trials (coin flips)

$\left(\frac{1}{2}\right)^x$  → the prob. of getting  $y$  heads in  $x$  flips, assuming a fair coin.

b

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

$$E(XY) = \sum_{x=1}^6 \sum_{y=0}^x xy \cdot f_{XY}(x, y)$$

we found  $f_{XY}(x, y) = \frac{1}{6} \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x$  for  $1 \leq x \leq 6$

and  $0 \leq y \leq x$ , we can compute  $E(XY)$ :

$$E(XY) = \sum_{x=1}^6 \sum_{y=0}^x xy \cdot \frac{1}{6} \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x \cdot \underbrace{\sum_{y=0}^x y \cdot \binom{x}{y} \cdot \left(\frac{1}{2}\right)^x}_{x \cdot 2^{x-1}}$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x \cdot x \cdot 2^{x-1} \cdot \left(\frac{1}{2}\right)^x$$

$$E(XY) = \frac{1}{6} \sum_{x=1}^6 x^2 \cdot \frac{1}{2}$$

$$E(XY) = \frac{1}{12} \sum_{x=1}^6 x^2$$

$$E(XY) = \frac{1}{12} (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)$$

$$E(XY) = \frac{91}{12}$$

$$E(x) = \sum_{x=1}^6 x \cdot P(X=x)$$

$$E(x) = \sum_{x=1}^6 x \cdot \frac{1}{6}$$

$$E(x) = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} \rightarrow E(x) = \frac{7}{2}$$

$$E(Y) = E(x \cdot p) = E(x) \cdot p = \frac{7}{2} \cdot 0.5 = \frac{7}{4} \rightarrow E(Y) = \frac{7}{4}$$

$$\text{Cov}(x, Y) = E(XY) - E(x)E(Y)$$

$$\text{Cov}(x, Y) = \frac{91}{12} - \frac{7}{2} \cdot \frac{7}{4} = \frac{91}{12} - \frac{49}{8} = \frac{35}{24}$$

$$\text{Cov}(x, Y) = \frac{35}{24}$$

---

(c)

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

$$\text{Var}(Y) = \sum_{x=1}^6 x \cdot p \cdot (1-p) = 21 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{21}{4}$$

$$\text{we found } \text{Cov}(X, Y) = \frac{35}{24}$$

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

$$E(X^2) = \sum_{x=1}^6 x^2 \cdot P(X=x) = \sum_{x=1}^6 x^2 \cdot \frac{1}{6} = \frac{91}{6}$$

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{182 - 147}{12} = \frac{35}{12}$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\frac{35}{24}}{\sqrt{\frac{35}{12} \cdot \frac{21}{4}}} = \frac{\frac{35}{24}}{\frac{7\sqrt{5}}{2}} =$$

$$= \frac{35}{24} \cdot \frac{2}{7\sqrt{5}} = \frac{5}{12\sqrt{5}} = \frac{5\sqrt{5}}{12 \cdot 5} = \frac{\sqrt{5}}{12}$$

$$\text{Corr}(X, Y) = \frac{\sqrt{5}}{12}$$

(d)

- The positive value indicates a tendency for both variables X and Y to move in the same direction. As one variable increases, the other tends to increase and vice versa
- However, the strength of this relation is not particularly strong. There is a discernible trend that the variables move together, but the relationship is not extremely strong
- Correlation does not mean causation. Even though these variables might show a relationship, it does not mean that changes in one variable cause changes in the other

In summary,  $\text{Corr}(X, Y) = \frac{\sqrt{5}}{12}$  indicates a moderate

positive linear relation between X and Y, suggesting that X and Y tend to move together in the same direction but not very strongly

### Question 3

(a)

Finding all the values of X and Y=3 according to the given joint probability:

$$f_{X,Y}(x|y) = \frac{f_{X,Y}(x,y)}{P(Y=y)}$$

$$P(Y=3) = f_{X,Y}(2,3) + f_{X,Y}(3,3) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

Computing the conditional probability distribution for X given that Y=3:

$$f_{X|Y=3}(x|3) = \frac{f_{X,Y}(x,3)}{P(Y=3)}$$

$$f_{X|Y=3}(2|3) = \frac{f_{X,Y}(2,3)}{P(Y=3)} = \frac{\frac{1}{5}}{\frac{2}{5}} = \frac{1}{2}$$

$$f_{X|Y=3}(3|3) = \frac{f_{X,Y}(3,3)}{P(Y=3)} = \frac{\frac{1}{5}}{\frac{2}{5}} = \frac{1}{2}$$

Calculating the conditional expectation of X given Y=3:

$$E[X|Y=3] = \sum_x x \cdot f_{X|Y=3}(x|3)$$

$$E[X|Y=3] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 1 + 1.5 = 2.5$$

$$E[X|Y=3] = 2.5$$

---

(b)

Finding all the values of Y and X=3 according to  
the given joint probability:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{P(X=x)}$$

$$P(X=3) = f_{X,Y}(3,2) + f_{X,Y}(3,3) + f_{X,Y}(3,17) =$$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$$

Computing the conditional probability distribution  
for Y given that X=3:

$$f_{Y|X=3}(y|3) = \frac{f_{X,Y}(3,y)}{P(X=3)}$$

$$f_{Y|X=3}(2|3) = \frac{f_{X,Y}(3,2)}{P(X=3)} = \frac{1/5}{3/5} = \frac{1}{3}$$

$$f_{Y|X=3}(3|3) = \frac{f_{X,Y}(3,3)}{P(X=3)} = \frac{1/5}{3/5} = \frac{1}{3}$$

$$f_{Y|X=3}(17|3) = \frac{f_{X,Y}(3,17)}{P(X=3)} = \frac{1/5}{3/5} = \frac{1}{3}$$

Calculating the conditional expectation of X given Y=3:

$$E[Y|X=3] = \sum_y y \cdot f_{Y|X=3}(y|3)$$

$$E[Y|X=3] = 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 17 \cdot \frac{1}{3} = \frac{2+3+17}{3} = \frac{22}{3}$$

$$E[Y|X=3] = \frac{22}{3}$$

(C)

$$P(Y=2) = f_{X|Y}(3|2) + f_{X|Y}(2,2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(Y=3) = f_{X|Y}(2,3) + f_{X|Y}(3,3) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(Y=17) = f_{X|Y}(3,17) = \frac{1}{5}$$

$$E[X|Y=2] = \sum_x x \cdot f_{X|Y=2}(x|2)$$

$$E[X|Y=2] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 1 + 1.5 = 2.5$$

$$E[X|Y=2] = 2.5$$

$$E[X|Y=3] = 2.5$$

$$E[X|Y=17] = \sum_x x \cdot f_{X|Y=17}(x|17)$$

$$E[X|Y=17] = 1 \cdot 3 = 3$$

$$E[X|Y=17] = 3$$

$$E[X|Y] = 2.5 \cdot \frac{2}{5} + 2.5 \cdot \frac{2}{5} + 3 \cdot \frac{1}{5} = \frac{13}{5} = 2.6$$

$$E[X|Y] = 2.6$$

(d)

$$P(X=2) = f_{X|Y}(2|3) + f_{X|Y}(2,2) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(X=3) = f_{X|Y}(3,2) + f_{X|Y}(3,3) + f_{X|Y}(3,1) =$$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}$$

$$E[Y|X=2] = \sum_y y \cdot f_{Y|X=2}(y|2)$$

$$E[Y|X=2] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = \frac{5}{2} = 2.5$$

$$E[Y|X=3] = \frac{22}{3}$$

$$E[Y|X] = \frac{5}{2} + \frac{22}{3} = \frac{15+44}{6} = \frac{59}{6}$$

## Question 4

$X_i \rightarrow$  the value shown on the  $i^{\text{th}}$  die when rolled

$$Z_n = X_1^2 + X_2^2 + \dots + X_n^2$$

Each  $X_i$  follows a discrete uniform distribution over  $\{1, 2, 3, 4, 5, 6\}$ , because we are rolling a fair 6-sided die

The mean of  $X_i^2$  is:

$$E[X_i^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$$

$$E[X_i^2] = \frac{1}{6} (1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

$Y_i = X_i^2$  for each  $i$ .

According to the strong law of large numbers, if  $Y_1, Y_2, \dots, Y_n$  are independent and identically distributed random variables with finite mean  $\mu = \frac{91}{6}$ , then:

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \mu$$

$$\text{In this case, } \frac{1}{n} Z_n = \frac{1}{n} (X_1^2 + X_2^2 + \dots + X_n^2) = \frac{1}{n} \sum_{i=1}^n Y_i$$

converges almost surely to  $\frac{91}{6}$  by the strong law of large numbers

So we have  $m = \frac{91}{6}$  as the limit of  $\frac{1}{n} Z_n$  as  $n$  approaches infinity

## Question 5

(a)

$$\lambda = \frac{1}{2} \text{ is given}$$

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2 \rightarrow \text{The standard deviation of the service time for a single customer}$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{16}} = \frac{2}{4} = 0.5 \text{ minutes}$$

$$X = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = 0.5 \text{ minutes}$$

$$Z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{0.5} = \frac{0.5}{0.5} = 1$$

$P(Z < 1) \rightarrow$  From the standard normal distribution table,  $P(Z < 1) = 0.8413$

b

$$\lambda = \frac{1}{2} \text{ is given}$$

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2 \rightarrow \text{The standard deviation of the service time for a single customer}$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{36}} = \frac{2}{6} = \frac{1}{3} \text{ minutes}$$

$$X = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{1}{3} \text{ minutes}$$

$$Z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{\frac{1}{3}} = \frac{0.5}{\frac{1}{3}} = 1.5$$

$P(Z < 1.5) \rightarrow$  From the standard normal distribution table,  $P(Z < 1.5) = 0.9332$

(c)

$$\lambda = \frac{1}{2} \text{ is given}$$

$$\mu = \frac{1}{\lambda} \rightarrow \mu = \frac{1}{\frac{1}{2}} = 2 \text{ minutes}$$

$$\sigma = \frac{1}{\lambda} \rightarrow \sigma = \frac{1}{\frac{1}{2}} = 2 \rightarrow \text{The standard deviation of the service time for a single customer}$$

$$\mu_{\bar{x}} = \mu = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{100}} = \frac{2}{10} = \frac{1}{5} \text{ minutes}$$

$$X = 2.5 \text{ minutes}$$

$$\mu_{\bar{x}} = 2 \text{ minutes}$$

$$\sigma_{\bar{x}} = \frac{1}{5} \text{ minutes}$$

$$Z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \rightarrow Z = \frac{2.5 - 2}{\frac{1}{5}} = \frac{0.5}{\frac{1}{5}} = 2.5$$

$P(Z < 2.5) \rightarrow$  From the standard normal distribution table,  $P(Z < 2.5) = 0.9938$

(d)

The CLT suggests that as  $n$  increases:

1. The probability that the average service time of the first  $n$  customers is less than 2.5 is going to increase
2. The distribution of sample means tends to become more symmetric normal distribution. It occurs irrespective of the initial distribution of the population, demonstrating a more standardized and even shape over time
3. The variability of the sample mean distribution decreases. This decrease in standard error causes a narrower range around the true population mean.
4. Larger sample sizes enhance the accuracy of statistical conclusions. They lead to more precise estimations, narrower confidence intervals, and more reliable hypothesis testing, improving the overall accuracy of inferences drawn from the data