Question 1

(a) Since all probabilities must sum to 1:

$$\begin{aligned}\sum_x \sum_y P_{X,Y}(x,y) &= 1 \\ 0.12 + 0.18 + 0.10 + 0.28 + a + 0.15 + 0.08 + 0.06 + 0.03 &= 1 \\ 1.00 + a &= 1 \\ a &= 0\end{aligned}$$

(b) We marginalize the joint PMF to get the marginal PMFs:
Marginal PMF of X:

$$\begin{aligned}P_X(1) &= \sum_y P_{X,Y}(1,y) = 0.12 + 0.18 + 0.10 = 0.40 \\ P_X(2) &= \sum_y P_{X,Y}(2,y) = 0.28 + 0 + 0.15 = 0.43 \\ P_X(3) &= \sum_y P_{X,Y}(3,y) = 0.08 + 0.06 + 0.03 = 0.17\end{aligned}$$

So that

$$P_X(x) = \begin{cases} 0.40 & x = 1 \\ 0.43 & x = 2 \\ 0.17 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Marginal PMF of Y:

$$\begin{aligned}P_Y(0) &= \sum_x P_{X,Y}(x,0) = 0.12 + 0.28 + 0.08 = 0.48 \\ P_Y(1) &= \sum_x P_{X,Y}(x,1) = 0.18 + 0 + 0.06 = 0.24 \\ P_Y(2) &= \sum_x P_{X,Y}(x,2) = 0.10 + 0.15 + 0.03 = 0.28\end{aligned}$$

So that

$$P_Y(y) = \begin{cases} 0.48 & y = 0 \\ 0.24 & y = 1 \\ 0.28 & y = 2 \\ 0 & \text{otherwise} \end{cases}$$

(c) For independence, we need $P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)$ for all x, y.
Let's test the independence condition by checking $(x = 1, y = 0)$:

$$\text{LHS: } P_{X,Y}(1,0) = 0.12$$
$$\text{RHS: } P_X(1) \cdot P_Y(0) = 0.40 \times 0.48 = 0.192$$

Since LHS $\neq$ RHS, X and Y are **not independent**.

(d) By the definition of the joint CDF,

$$\begin{aligned}F_{X,Y}(2,1) &= P(X \leq 2, Y \leq 1) \\ &= P_{X,Y}(1,0) + P_{X,Y}(1,1) + P_{X,Y}(2,0) + P_{X,Y}(2,1) \\ &= 0.12 + 0.18 + 0.28 + 0 \\ &= 0.58.\end{aligned}$$

(e) The event $X + Y \leq 3$ contains the points

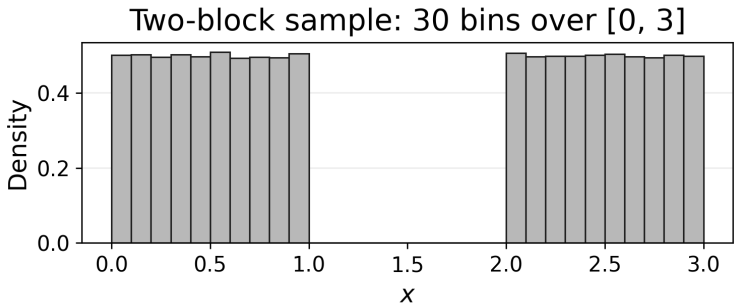
$$(1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (3, 0).$$

Therefore,

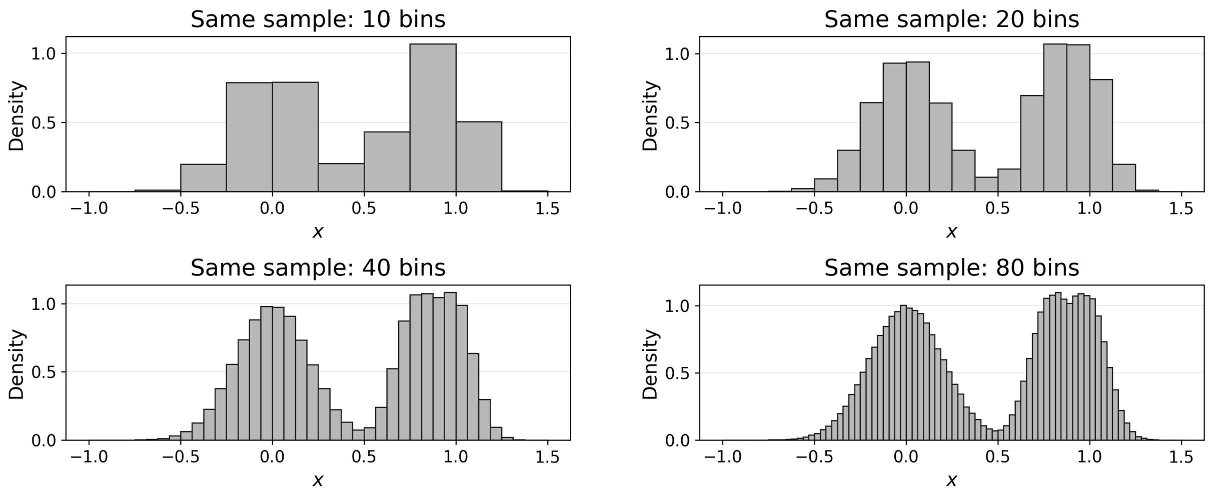
$$\begin{aligned} P(X + Y \leq 3) &= 0.12 + 0.18 + 0.10 + 0.28 + 0 + 0.08 \\ &= 0.76. \end{aligned}$$

Question 2

(a) The 30 equal-width bins span $[0, 3]$, so each bin has width 0.1. The density is zero on $[1, 2]$, which accounts for $1/0.1 = 10$ empty bins. The remaining 20 bins lie in $[0, 1]$ and $[2, 3]$ and are non-empty.



(b) Four histograms below are all from the same sample, using 10, 20, 40 and 80 bins.



With 10 bins, the estimate is very smooth but too coarse: the nearby components in the right-hand group are merged. At 20 bins, the two broad groups are clear, but their finer structure is still obscured. At 40 bins, more detail appears and the right-hand group begins to flatten or separate near its top. At 80 bins, the increased resolution reveals three component peaks in total: one near 0 and two nearby peaks around 0.8 and 1.0. Because the sample is large, the 80-bin histogram remains reasonably smooth here.

(c) For this sample, 80 bins is a defensible choice when the aim is to identify the distribution's components: it resolves the two nearby right-hand peaks without introducing severe bin-to-bin noise. A 40-bin display would also be reasonable if only the broad overall shape were required. In general, the choice balances smoothness against the resolution of fine features.

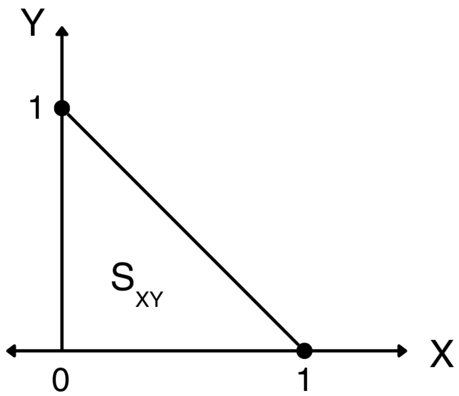
(d) For a density histogram, the height of bin j is its relative frequency divided by its width:

$$h_j = \frac{n_j}{N\Delta} = \frac{30}{200(0.25)} = 0.6.$$

As a check, the area of the bar is $h_j\Delta = 0.6(0.25) = 0.15 = n_j/N$, the observed proportion in the bin.

Question 3

(a) The range of X and Y is the triangle bounded by $x \geq 0$, $y \geq 0$, and $x + y < 1$.



(b) To find c:

$$\begin{aligned} 1 &= \int_0^1 \int_0^{1-x} (cy+1)\,dy\,dx \\ &= \int_0^1 \left[\tfrac{c}{2}(1-x)^2 + (1-x)\right] dx \\ &= \frac{c}{6} + \frac{1}{2}. \end{aligned}$$

Hence $c = 3$.

(c) For the marginals:

$$f_X(x) = \int_0^{1-x} (3y+1)\,dy = \left[\tfrac{3}{2}y^2 + y\right]_0^{1-x} = \tfrac{3}{2}(1-x)^2 + (1-x).$$

Thus

$$f_X(x) = \begin{cases} \tfrac{3}{2}(1-x)^2 + (1-x), & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

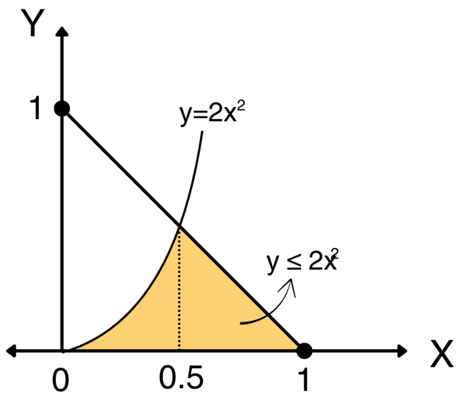
For $f_Y(y)$:

$$f_Y(y) = \int_0^{1-y} (3y+1)\,dx = (3y+1)(1-y).$$

So

$$f_Y(y) = \begin{cases} (3y+1)(1-y), & 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

(d) Compute $P(Y < 2X^2)$:



The curves $y = 2x^2$ and $y = 1 - x$ intersect at $x = 1/2$. Integrating with respect to x first, the event has bounds $0 \le y \le 1/2$ and $\sqrt{y/2} < x \le 1 - y$.

Therefore,

$$\begin{aligned}
P(Y < 2X^2) &= \int_0^{1/2} \int_{\sqrt{y/2}}^{1-y} (3y+1)\,dx\,dy \\
&= \int_0^{1/2} (3y+1)\left[(1-y) - \sqrt{\frac{y}{2}}\right] dy \\
&= \int_0^{1/2} (3y+1)(1-y)\,dy - \int_0^{1/2} (3y+1)\sqrt{\frac{y}{2}}\,dy \\
&= \int_0^{1/2} (2y+1-3y^2)\,dy - \frac{1}{\sqrt{2}} \int_0^{1/2} (3y^{3/2} + y^{1/2})\,dy \\
&= \left[y^2 + y - y^3\right]_0^{1/2} - \frac{1}{\sqrt{2}} \left[\frac{6y^{5/2}}{5} + \frac{2y^{3/2}}{3}\right]_0^{1/2} \\
&= \frac{5}{8} - \frac{1}{\sqrt{2}} \left[\frac{6}{5} \cdot \frac{1}{4\sqrt{2}} + \frac{2}{3} \cdot \frac{1}{2\sqrt{2}}\right] \\
&= \frac{5}{8} - \frac{1}{2}\left[\frac{3}{10} + \frac{1}{3}\right] \\
&= \frac{37}{120} \\
&= 0.308.
\end{aligned}$$

Question 4

(a) First verify this is a valid PDF:

$$\int_0^1 \int_0^x 2\,dy\,dx = \int_0^1 2x\,dx = \left[x^2\right]_0^1 = 1 \quad \checkmark$$

We calculate the expected values using the definition $\mathrm{E}[g(X,Y)] = \int\int g(x,y) f_{X,Y}(x,y)\,dx\,dy$:

$$\mathrm{E}[X] = \int_0^1 \int_0^x x \cdot 2\,dy\,dx = \int_0^1 2x^2\,dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$

$$\mathrm{E}[Y] = \int_0^1 \int_0^x y \cdot 2\,dy\,dx = \int_0^1 2\left[\frac{y^2}{2}\right]_0^x dx = \int_0^1 x^2\,dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

$$r_{X,Y} = \mathrm{E}[XY] = \int_0^1 \int_0^x xy \cdot 2\,dy\,dx = \int_0^1 2x\left[\frac{y^2}{2}\right]_0^x dx = \int_0^1 x^3\,dx = \left[\frac{x^4}{4}\right]_0^1 = \frac{1}{4}$$

(b) Using linearity of expectation:

$$\mathrm{E}[2X + 3Y] = 2\mathrm{E}[X] + 3\mathrm{E}[Y] = 2 \cdot \frac{2}{3} + 3 \cdot \frac{1}{3} = \frac{4}{3} + 1 = \frac{7}{3}$$

For $\mathrm{E}[X^2 + Y^2]$, we first calculate $\mathrm{E}[X^2]$ and $\mathrm{E}[Y^2]$:

$$\mathrm{E}[X^2] = \int_0^1 \int_0^x x^2 \cdot 2\,dy\,dx = \int_0^1 2x^3\,dx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1}{2}$$

$$\mathrm{E}[Y^2] = \int_0^1 \int_0^x y^2 \cdot 2\,dy\,dx = \int_0^1 2\left[\frac{y^3}{3}\right]_0^x dx = \int_0^1 \frac{2x^3}{3}\,dx = \left[\frac{x^4}{6}\right]_0^1 = \frac{1}{6}$$

Therefore: $\mathrm{E}[X^2 + Y^2] = \frac{1}{2} + \frac{1}{6} = \frac{3+1}{6} = \frac{2}{3}$

(c) Using the formula $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$:

$$\begin{aligned}\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] &= \frac{1}{4} - \frac{2}{3} \cdot \frac{1}{3} = \frac{1}{4} - \frac{2}{9} \\ &= \frac{9-8}{36} = \frac{1}{36}\end{aligned}$$

For correlation, we use $\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]} \cdot \sqrt{\mathrm{Var}[Y]}}$:

First calculate the variances using $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$:

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \frac{1}{18}$$

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{6} - \frac{1}{9} = \frac{3-2}{18} = \frac{1}{18}$$

Therefore:

$$\rho_{X,Y} = \frac{\frac{1}{36}}{\sqrt{\frac{1}{18}} \cdot \sqrt{\frac{1}{18}}} = \frac{\frac{1}{36}}{\frac{1}{18}} = \frac{1}{36} \cdot \frac{18}{1} = \frac{18}{36} = \frac{1}{2}$$

(d) Use

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y].$$

Substituting the values obtained above,

$$\begin{aligned}\mathrm{Var}[X+Y] &= \frac{1}{18} + \frac{1}{18} + 2\left(\frac{1}{36}\right) \\ &= \frac{1}{6}.\end{aligned}$$

(e) For affine rescalings $U = aX + b$ and $V = cY + d$,

$$\mathrm{Cov}[U,V] = ac\,\mathrm{Cov}[X,Y], \qquad \rho_{U,V} = \mathrm{sign}(ac)\rho_{X,Y}.$$

Here $a = 100$ and $c = -2$, so

$$\begin{aligned}\mathrm{Cov}[U,V] &= 100(-2)\left(\frac{1}{36}\right) = -\frac{50}{9}, \\ \rho_{U,V} &= \mathrm{sign}(-200)\left(\frac{1}{2}\right) = -\frac{1}{2}.\end{aligned}$$

Question 5

The joint PDF is

$$f_{X,Y}(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

It is non-negative and its total integral is

$$\int_0^1 \int_0^1 (x+y)\, dy\, dx = 1,$$

so it is a valid joint PDF.

(a) For $0 < x \le 1$ and $0 < y \le 1$,

$$\begin{aligned} F_{X,Y}(x,y) &= \int_0^x \int_0^y (u+v)\, dv\, du \\ &= \frac{x^2 y}{2} + \frac{xy^2}{2} = \frac{xy(x+y)}{2}. \end{aligned}$$

Accounting for values outside the unit-square support gives

$$F_{X,Y}(x,y) = \begin{cases} 0, & x \le 0 \text{ or } y \le 0, \\ \dfrac{xy(x+y)}{2}, & 0 < x \le 1,\ 0 < y \le 1, \\ \dfrac{y(1+y)}{2}, & x > 1,\ 0 < y \le 1, \\ \dfrac{x(1+x)}{2}, & 0 < x \le 1,\ y > 1, \\ 1, & x > 1,\ y > 1. \end{cases}$$

(b) Marginalise the joint PDF:

$$\begin{aligned} f_X(x) &= \int_0^1 (x+y)\, dy = x + \frac{1}{2}, & 0 \le x \le 1, \\ f_Y(y) &= \int_0^1 (x+y)\, dx = y + \frac{1}{2}, & 0 \le y \le 1. \end{aligned}$$

Thus each marginal is zero outside $[0,1]$.

(c) The four-corner formula is

$$\begin{aligned} P(x_1 < X \le x_2, y_1 < Y \le y_2) = F_{X,Y}(x_2,y_2) &- F_{X,Y}(x_1,y_2) \\ - F_{X,Y}(x_2,y_1) &+ F_{X,Y}(x_1,y_1). \end{aligned}$$

Therefore,

$$\begin{aligned} P\left(0 < X \le \tfrac{1}{2}, \tfrac{1}{4} < Y \le \tfrac{3}{4}\right) &= F_{X,Y}\left(\tfrac{1}{2}, \tfrac{3}{4}\right) - F_{X,Y}\left(0, \tfrac{3}{4}\right) \\ &\quad - F_{X,Y}\left(\tfrac{1}{2}, \tfrac{1}{4}\right) + F_{X,Y}\left(0, \tfrac{1}{4}\right) \\ &= \frac{15}{64} - 0 - \frac{3}{64} + 0 \\ &= \frac{3}{16}. \end{aligned}$$

(d) Independence would require

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

throughout the support. At $x = y = \frac{1}{4}$,

$$f_{X,Y}\left(\tfrac{1}{4}, \tfrac{1}{4}\right) = \frac{1}{2}, \qquad f_X\left(\tfrac{1}{4}\right) f_Y\left(\tfrac{1}{4}\right) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}.$$

These are unequal, so X and Y are **not independent**.

Question 6

Let X have a discrete uniform distribution with support $S_X = \{-1, 0, 1\}$, and let $Y = X^2$.

(a) The deterministic relation $Y = X^2$ gives

$P_{X,Y}(x,y)$	$y=0$	$y=1$
$x=-1$	0	$\frac{1}{3}$
$x=0$	$\frac{1}{3}$	0
$x=1$	0	$\frac{1}{3}$

(b) Use

$$\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

The required moments are

$$\begin{aligned}\mathrm{E}[X] &= \frac{-1+0+1}{3} = 0,\\ \mathrm{E}[Y] &= \frac{1+0+1}{3} = \frac{2}{3},\\ \mathrm{E}[XY] = \mathrm{E}[X^3] &= \frac{-1+0+1}{3} = 0.\end{aligned}$$

Hence

$$\mathrm{Cov}[X,Y] = 0 - 0\left(\frac{2}{3}\right) = 0,$$

so X and Y are uncorrelated.

(c) Independence would require every joint probability to factor into its marginals. However,

$$P(X=0, Y=0) = \frac{1}{3}, \qquad P_X(0)P_Y(0) = \frac{1}{3}\cdot\frac{1}{3} = \frac{1}{9}.$$

Therefore X and Y are **dependent**, even though their covariance is zero. This demonstrates that uncorrelated random variables need not be independent.