Lecture 7: Joint Distributions

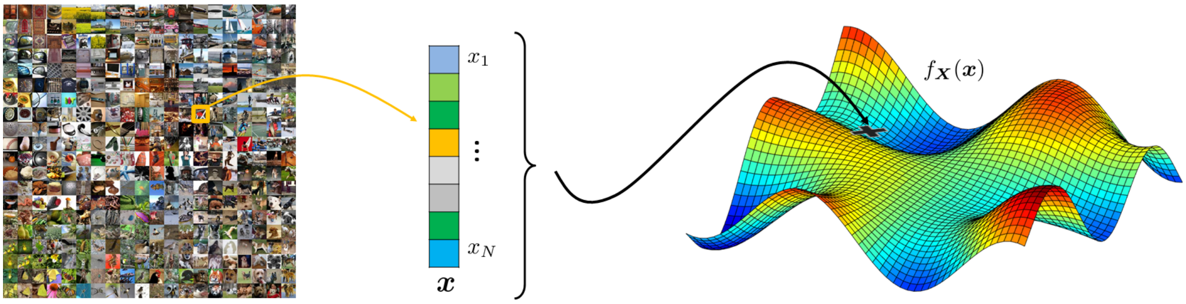
Yunwen Lei

School of Computer Science, University of Birmingham

1 Joint Distributions

In the previous class, we consider single random variables. In this class, we will consider the interaction between several random variables. A key concept to this aim is the joint distribution. Intuitively speaking, joint distributions are multi-dimensional PDF (or PMF or CDF)

$$\underbrace{f_X(x)}_{\text{one variable}} \Rightarrow \underbrace{f_{X_1,X_2}(x_1,x_2)}_{\text{two variables}} \Rightarrow \underbrace{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}_{\text{three variables}} \Rightarrow \ldots \Rightarrow \underbrace{f_{X_1,\ldots,X_N}(x_1,\ldots,x_N)}_{N \text{ variables}}$$

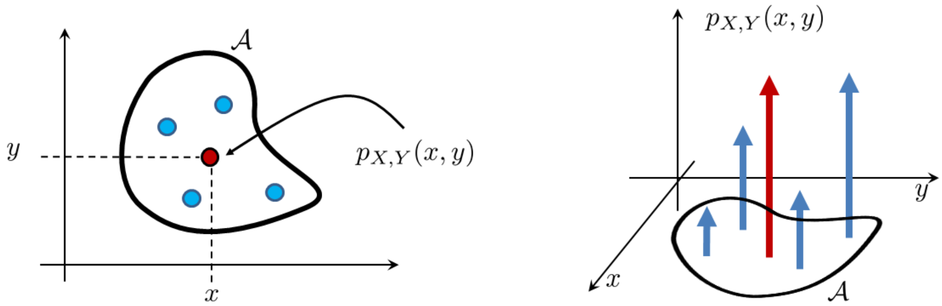


Joint distributions are ubiquitous in modern data analysis. For example, an image from a dataset can be represented by a high-dimensional vector $\mathbf{x} \in \mathbb{R}^d$, where each component x_i is random. Each vector has certain probability to be present. Such probability is described by the high-dimensional joint PDF $f_{\mathbf{X}}(\mathbf{x})$.

Definition 1 (Joint PMF)**.** Let X and Y be two discrete random variables. The joint PMF of X and Y is defined as

$$P_{X,Y}(x,y) = \mathbb{P}(X = x \text{ and } Y = y). \tag{1.1}$$

According to the definition, we know that joint PMF indicates the probability of the intersection of events $\{\xi : X(\xi) = x\}$ and $\{\xi : Y(\xi) = y\}$. Notice this definition requires both X and Y to be discrete random variables.



A joint PMF for a pair of discrete random variables consists of an array of impulses. To measure the size of the event $\mathcal{A}$, we sum all the impulses inside $\mathcal{A}$.

Example 1. We consider the experiment of tossing a coin and a dice. Let X be the coin toss, and Y be the dice.

Question. Find the joint PMF.

Solution. The sample space of X is $\{0, 1\}$. The sample space of Y is $\{1, 2, 3, 4, 5, 6\}$. The joint PMF is

	Y					
	1	2	3	4	5	6
$X = 0$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
$X = 1$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

Question. Define $\mathcal{A} = \{X + Y = 4\}$. Find $\mathbb{P}(\mathcal{A})$.

$$\textbf{Solution} : \mathbb{P}(A) = \sum_{(x,y)\in\mathcal{A}} P_{X,Y}(x, y) = P_{X,Y}(0, 4) + P_{X,Y}(1, 3) = \frac{2}{12}.$$

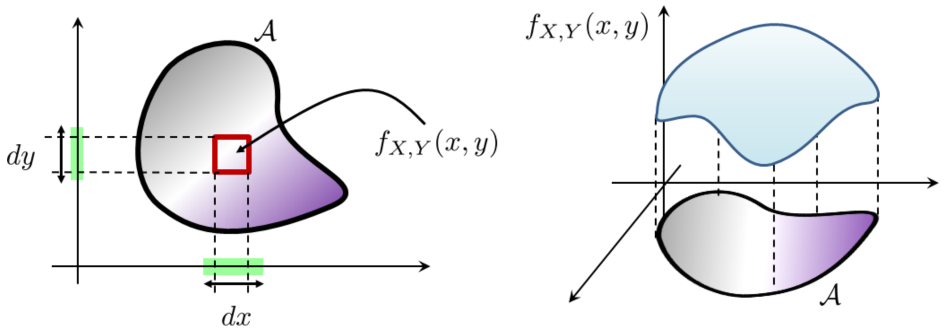
Similarly, we can also define the joint PDF for continuous random variables.

Definition 2 (Joint PDF)**.** Let X and Y be two continuous random variables. The joint PDF of X and Y is a function $f_{X,Y}(x, y)$ that can be integrated to yield a probability:

$$\mathbb{P}(\mathcal{A}) = \int_{\mathcal{A}} f_{X,Y}(x, y)dxdy \tag{1.2}$$

for any event $\mathcal{A} \subseteq X(\Omega) \times Y(\Omega)$.

This definition is a direct extension from one random variable to two random variables. The joint PDF can be interpreted as the probability per unit, and can take values larger than 1.



We now provide some examples to compute probability from PDF.

Example 2. Consider a uniform joint PDF $f_{X,Y}(x, y)$ defined on $[0, 2] \times [0, 2]$ with $f_{X,Y}(x, y) = \frac{1}{4}$.

Question 1. Let $\mathcal{A} = [a, b] \times [c, d]$. Find $\mathbb{P}(A)$.

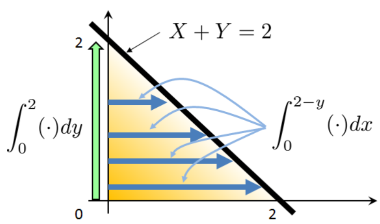
Solution:

$$\begin{aligned}\mathbb{P}(\mathcal{A}) &= \mathbb{P}\big(a \le X \le b, c \le Y \le d\big) \\ &= \int_c^d \int_a^b f_{X,Y}(x, y)dxdy = \int_c^d \int_a^b \frac{1}{4}dxdy = \frac{(d-c)(b-a)}{4}\end{aligned}$$

Question 2. Let $\mathcal{B} = \{X + Y \le 2\}$. Find $\mathbb{P}(\mathcal{B})$.

Solution:

$$\begin{aligned}\mathbb{P}(\mathcal{B}) = \int_{\mathcal{B}} f_{X,Y}(x, y) &= \int_0^2 \int_0^{2-y} f_{X,Y}(x, y)dxdy \\ &= \int_0^2 \int_0^{2-y} \frac{1}{4}dxdy \\ &= \int_0^2 \frac{2-y}{4}dy = \frac{1}{2}.\end{aligned}$$



The joint distribution shows how the two random variables are correlated. We can derive from joint distributions the marginal distributions for a single variable. This can be useful if we are only interested in a single random variable. For discrete random variables, we get marginal PMF by taking a summation over the random variable we are not interested in. For continuous random variables, we get marginal PDF by taking an integral over the random variable we are not interested in.

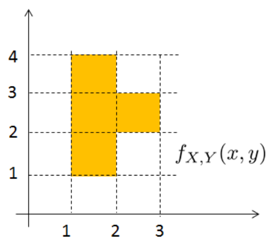
Definition 3 (Marginal PMF and Marginal PDF)**.** The marginal PMF is defined as

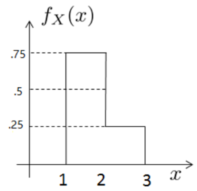
$$P_X(x) = \sum_{y \in Y(\Omega)} P_{X,Y}(x, y) \quad \text{and} \quad P_Y(y) = \sum_{x \in X(\Omega)} P_{X,Y}(x, y).$$

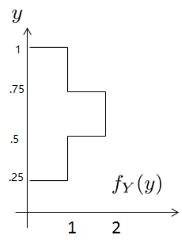
The marginal PDF is defined as

$$f_X(x) = \int_{Y(\Omega)} f_{X,Y}(x, y)dy \quad \text{and} \quad f_Y(y) = \int_{X(\Omega)} f_{X,Y}(x, y)dx.$$

Example 3. Consider the joint PDF $f_{X,Y}(x, y) = 1/4$ shown below. Then the PDF is $1/4$ for each point in the yellow region, and 0 otherwise. We can use this joint PDF to compute marginal PDFs.







We can compute $f_X(1)$ as follows

$$\begin{aligned} f_X(1) &= \int_{y \in Y(\Omega)} f_{X,Y}(1, y)dy \\ &= \int_1^4 f_{X,Y}(1, y)dy \\ &= \int_1^4 \frac{1}{4}dy = \frac{3}{4}. \end{aligned}$$

The PDF of f_X can be computed by

$$f_X(x) = \begin{cases} 3/4, & \text{if } x \in (1, 2] \\ 1/4, & \text{if } x \in (2, 3] \\ 0, & \text{otherwise.} \end{cases}$$

In a similar way, we can compute $f_Y(y)$ as follows

$$f_Y(y) = \begin{cases} 1/4, & \text{if } y \in (1, 2] \\ 2/4, & \text{if } y \in (2, 3] \\ 1/4, & \text{if } y \in (3, 4] \\ 0, & \text{otherwise.} \end{cases}$$

We can give the definition of joint CDF for two random variables. This definition is an extension of CDF to two random variables.

Definition 4 (Joint CDF)**.** Let X and Y be two random variables. The joint CDF of X and Y is the function $F_{X,Y}(x, y)$ such that

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x \cap Y \leq y).$$

In the following theorem, we show how to compute the joint CDF once we are given either joint PMF or joint PDF.

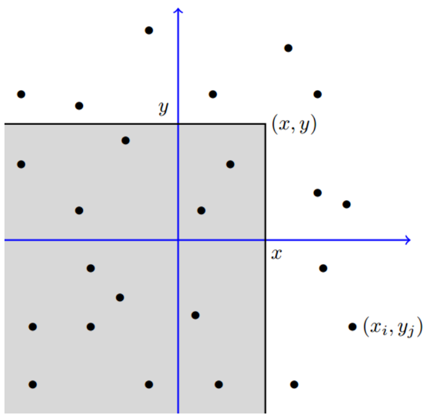
Theorem 1. *If X and Y are discrete, then*

$$F_{X,Y}(x, y) = \sum_{y' \leq y} \sum_{x' \leq x} P_{X,Y}(x', y'). \tag{1.3}$$

If X and Y are continuous, then

$$F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(x', y')dx'dy'. \tag{1.4}$$

We now give a geometric understanding of joint CDF. The dots represent the pairs $(x_i, y_j) \in X(\Omega) \times Y(\Omega)$. $F_{X,Y}(x, y)$ is the probability that (X, Y) belongs to the shaded region.



We can use joint CDF to define independency. Intuitively, we say two random variables are independent iff the joint CDF can be factorized into CDFs of single random variables. This factorization shows no correlation between these two random variables.

Definition 5 (Independency)**.** We say two random variables X and Y are independent if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

2 Conditional PMF and Conditional PDF

Analogous to conditional probability for events, we can define the conditional probability for random variables. We first consider the conditional PMF of a discrete random variable under the condition of an event. Notice this definition is consistent with that for events.

Definition 6 (Conditional PMF)**.** For a discrete random variable X and event A, the conditional PMF of X given A is defined as

$$\begin{aligned} P_{X|A}(x_i) &= \mathbb{P}(X = x_i|A) \\ &= \frac{\mathbb{P}(X = x_i \text{ and } A)}{\mathbb{P}(A)}, \quad \text{for any } x_i \in X(\Omega). \end{aligned} \tag{2.1}$$

Example 4. Let's roll a fair six-sided die. Let X be the observed number. Find the conditional PMF of X given the event A that the observed number was less than 5.

Answer: The PMF is conditioned on the event $A = \{X < 5\}$, where $\mathbb{P}(A) = \mathbb{P}(X < 5) = \frac{4}{6}$. Thus,

$$P_{X|A}(1) = \mathbb{P}(X = 1|X < 5) = \frac{\mathbb{P}(X = 1 \text{ and } X < 5)}{\mathbb{P}(X < 5)} = \frac{\mathbb{P}(X = 1)}{\mathbb{P}(X < 5)} = \frac{\frac{1}{6}}{\frac{4}{6}} = \frac{1}{4}.$$

Similarly, we can obtain

$$P_{X|A}(2) = P_{X|A}(3) = P_{X|A}(4) = \frac{1}{4}$$

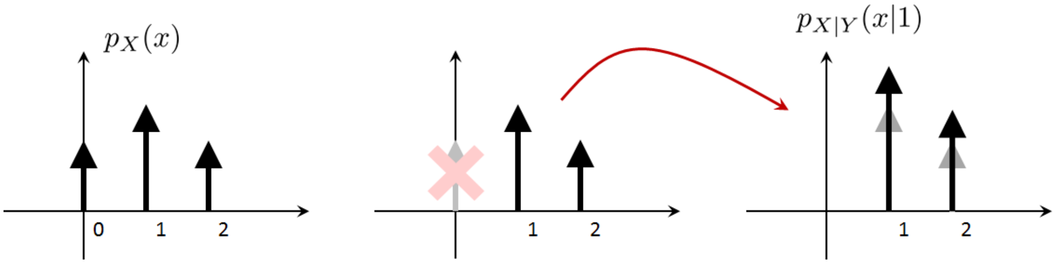
Since $\{X = 5\}$, $\{X = 6\}$ and event A are disjoint, we know $P_{X|A}(5) = P_{X|A}(6) = 0$.

The above definition of conditional PMF conditioned on an event can be generalised to random variables as follows

Definition 7 (Conditional PMF)**.** Let X and Y be two discrete random variables. The conditional PMF of X given Y is

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x, y)}{P_Y(y)}.$$

According to the definition, the conditional PMF is the division of the joint PMF and the marginal PMF.



Suppose X is the sum of two coins, and Y is the first coin. When X is unconditioned, the PMF is just $P_X(x)$. When X is conditioned on $Y = 1$, then $X = 0$ cannot happen. Therefore, the resulting PMF $P_{X|Y}(x|1)$ only has two states. After normalization we obtain the conditional PMF.

Example 5. Consider a joint PMF given in the following table. Find the conditional PMF $P_{X|Y}(x|1)$ and the marginal PMF $P_X(x)$

	Y=			
	1	2	3	4
$X = 1$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{0}{20}$
2	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
3	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
4	$\frac{0}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$

Solution: We can take a summation over different rows to get the marginal PMF $P_X(x)$

$$P_X(x) = \left(\frac{3}{20}, \frac{7}{20}, \frac{7}{20}, \frac{3}{20}\right).$$

To compute $P_{X|Y}(x|1)$, we only need to consider the column corresponding to $Y = 1$. This corresponds to a normalization of that column: divide each term by $P_Y(1)$

$$P_{X|Y}(x|1) = \frac{P_{X,Y}(x,1)}{P_Y(1)} = \frac{\left(\frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{0}{20}\right)}{\frac{3}{20}} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0\right)$$

If we consider continuous random variables, we can define the conditional PDF.

Definition 8 (Conditional PDF)**.** Let X and Y be two continuous random variables. The conditional PDF of X given Y is

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

The conditional PDF to be interpreted as

$$f_{X|Y}(x|y)\delta_x \approx \mathbb{P}(x < X \leq x + \delta_x | Y = y).$$

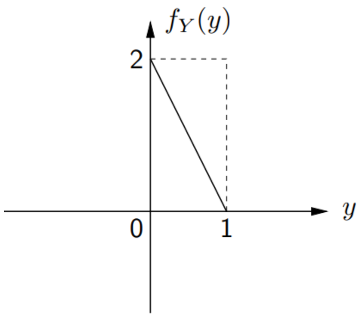
That is, the conditional PDF can be interpreted as the conditional probability per unit. Therefore, the conditional PDF can take values larger than 1. This is consistent with the fact that PDF can be interpreted as the probability per unit.

Example 6. We now provide an example to compute conditional PDF when we are given joint PDF.

$$\text{Let} \quad f(x,y) = \begin{cases} 2, & \text{if } x, y \geq 0, x + y \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

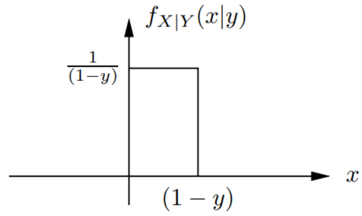
We can take an integral over x to get the marginal PDF for Y

$$f_Y(y) = \int_0^1 f(x,y)dx = \begin{cases} \int_0^{1-y} 2dx, & \text{if } y \in [0,1] \\ 0, & \text{otherwise.} \end{cases}.$$



Therefore, we can compute the conditional PDF as follows

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1-y}, & \text{if } x, y \geq 0, x + y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$



We can use PMF and PDF to define independency for two random variables. Two random variables are independent if the joint PMF/PDF can be factored into PMF/PDF for each random variable. This definition is equivalent to Definition 5, which applies to both continuous and discrete random variables.

Definition 9 (Independence for Two Variables)**.** We say two discrete random variables X and Y are independent iff

$$P_{X,Y}(x,y) = P_X(x)P_Y(y). \tag{2.3}$$

We say two continuous random variables X and Y are independent iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y). \tag{2.4}$$

We can generalize this definition to the case with more than two random variables. For simplicity, we consider continuous random variables. The definition for discrete random variables is similar except replacing the joint PDF with the joint PMF.

Definition 10 (Independence for Multiple Variables)**.** We say a sequence of random variables $X_1, X_2, \ldots, X_N$ are independent iff the joint PDF can be factorized

$$f_{X_1,\ldots,X_N}(x_1,\ldots,x_N) = \prod_{n=1}^{N} f_{X_n}(x_n). \tag{2.5}$$

Below we provide two examples to show how to verify two random variables are independent or not.

Example 7. Consider two random variables with a joint PDF given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{(x-\mu_X)^2+(y-\mu_Y)^2}{2\sigma^2}\right\}. \tag{2.6}$$

Then, we can decompose the joint PDF as follows

$$f_{X,Y}(x,y) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu_X)^2}{2\sigma^2}\right\}}_{f_X(x)}\underbrace{\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma^2}\right\}}_{f_Y(y)}$$

This shows that we can factorize the joint PDF into a function of X and a function of Y (no interacting term). Therefore, the random variables X and Y are independent.

Example 8. Consider two random variables X and Y with a joint PDF given by

$$\begin{aligned} f_{X,Y}(x,y) &\propto \exp\left\{-(x-y)^2\right\} = \exp\left\{-x^2+2xy-y^2\right\} \\ &= \underbrace{\exp\{-x^2\}}_{f_X(x)}\underbrace{\exp\{2xy\}}_{\text{extra term}}\underbrace{\exp\{-y^2\}}_{f_Y(y)} \end{aligned}$$

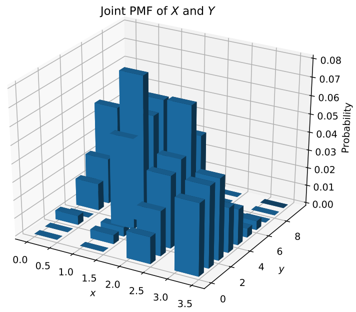
There is a term showing the interaction between X and Y. This PDF cannot be factorized into a product of two marginal PDFs. Therefore, X and Y are dependent.

3 Example

Suppose we are given two random variables X and Y with joint PMF in the following table

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$	$P_X(x)$
$x=0$	$\frac{4}{24}$	$\frac{6}{24}$	$\frac{3}{24}$	
$x=1$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{4}{24}$	
$P_Y(y)$				

Joint PMF of X and Y and their marginal PMFs $P_X(x)$ and $P_Y(y)$



Questions:

- Find the marginal PMFs of X and Y and complete the table
- Find $\mathbb{P}(X=0, Y \leq 1)$.
- Find $\mathbb{P}(Y=1 \,|\, X=0)$.
- Are X and Y independent?

We now solve these questions one by one.

A1: To find the marginal PMFs of X and Y and complete the table, we first need to obtain the ranges of X and Y from the table: $X(\Omega) = \{0,1\}$ and $Y(\Omega) = \{0,1,2\}$. Then we compute the marginal PMFs:

$$P_X(0) = \sum_{y_i \in Y(\Omega)} P_{X,Y}(X=0, y_i) = \frac{4}{24} + \frac{6}{24} + \frac{3}{24} = \frac{13}{24},$$

that is, we sum the values of the three columns in the row of $X=0$ to obtain the probability of $P_X(0)$. Similarly, we can obtain $P_X(1) = \frac{11}{24}$.

To summarise:

$$P_X(x) = \begin{cases} \frac{13}{24} & x=0 \\ \frac{11}{24} & x=1 \\ 0 & \text{otherwise} \end{cases}$$

In a similar way, we have

$$P_Y(0) = \sum_{x_j \in X(\Omega)} P_{X,Y}(x_j, Y=0) = \frac{4}{24} + \frac{3}{24} = \frac{7}{24}.$$

That is, we sum the value in the column of $Y=0$ to obtain the probability of $P_Y(0)$. Similarly, we can obtain $P_Y(1) = \frac{10}{24}$ and $P_Y(2) = \frac{7}{24}$. To summarise:

$$P_Y(y) = \begin{cases} \frac{7}{24} & y=0 \\ \frac{10}{24} & y=1 \\ \frac{7}{24} & y=2 \\ 0 & \text{otherwise} \end{cases}$$

Now we can complete the table.

	$Y=0$	$Y=1$	$Y=2$	$P_X(x)$
$X=0$	$\frac{4}{24}$	$\frac{6}{24}$	$\frac{3}{24}$	$\frac{13}{24}$
$X=1$	$\frac{3}{24}$	$\frac{4}{24}$	$\frac{4}{24}$	$\frac{11}{24}$
$P_Y(y)$	$\frac{7}{24}$	$\frac{10}{24}$	$\frac{7}{24}$	$\frac{24}{24}$

Table 1: Joint PMF of X and Y and their marginal PMFs $P_X(x)$ and $P_Y(y)$

Q2: Find $\mathbb{P}(X = 0, Y \leq 1)$.

A2: To find $\mathbb{P}(X = 0, Y \leq 1)$, we can write

$$\mathbb{P}(X = 0, Y \leq 1) = P_{X,Y}(0,0) + P_{X,Y}(0,1) = \frac{4}{24} + \frac{6}{24} = \frac{10}{24}.$$

Q3: Find $\mathbb{P}(Y = 1 \,|\, X = 0)$.

A3: Using the formula for conditional probability, we have

$$\mathbb{P}(Y = 1 \,|\, X = 0) = \frac{\mathbb{P}(Y = 1, X = 0)}{\mathbb{P}(X = 0)} = \frac{P_{X,Y}(0,1)}{P_X(0)} = \frac{6}{13}$$

Q4: Are X and Y independent?

A4: Independent means $\mathbb{P}(A|B) = \mathbb{P}(A)$, however, a counter example as we just derived

$$\mathbb{P}(Y = 1 \,|\, X = 0) = \frac{6}{13} \neq \mathbb{P}(Y = 1) = \frac{10}{24}$$

can prove they are not independent.