# WEEK 3

## 3. Random Variables and Their Distributions

As it is mentioned before, statistics is a science of collecting, organizing and interpreting the numerical facts which we called data. Collection of data requires an experiment. That is, we collect the data by experiment and we repeat the experiment many times (as much as we can) we observe a set of data. However, when you perform an experiment, the results of the outcomes are not numerical values. For example, when you toss a coin, you observe either a head or a tail. These outcomes are not numerical values. However, if we transfer these outcomes to real numbers (for example, when you observe a head you can match to zero and when you observe a tail you can match to one). Then you can do the calculations with zeros and ones. Therefore, after the experiment, first we need to transfer the outcomes to numerical values. The random variable makes such a transformation.

**Definition**: A random variable, is a function from the sample space to the real numbers, namely,

$$X:\Omega \to \mathbb{R}$$
$$w \to X(w).$$

The range of a random variable (say $D_X$ ) is a subset of real line. That is, $D_X \subset \mathbb{R}$. The subsets of the real numbers could be either a countable or uncountable subsets.

Figure 3.1

If, $D_X$ is a countable subset of $\mathbb{R}$ then the random variable is called a discrete random variable otherwise it is a continuous random variable.

**Example**: a) Consider an experiment of tossing a coin three times. Then the sample space (which is a set of all possible outcomes) is $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. Define a function $X$ as follow:

$$X : \Omega \to \mathbb{R}$$

$$w \to X(w) = \begin{cases} 0 & , \quad w = HHH \\ 1 & , \quad w = HHT, HTH, THH \\ 2 & , \quad w = TTH, THT, HTT \\ 3 & , \quad w = TTT \end{cases}$$

Here, the function $X$ counts the number of tails in the experiment. The range of the function $X$ is $D_X = \{0,1,2,3\}$. This subset is a countable subset of the real line. Thus, the random variable is discrete.

b) Consider an experiment of the weights of new-born-babies. Here, the sample space is all new-born-babies in a hospital. Define a random variable

$$X : \Omega \to \mathbb{R}$$

$$w \to X(w).$$

The weights of the baby could be any real number (say between 2.5 kg to 5 kg). So, the range of the random variable is $D_X = (2.5, 5)$. This is an uncountable subset of the real numbers. Therefore, the random variable is continuous.

c) Consider an experiment of student's performance in an examination. Students may get any score between $0$ to $100$. That is, the range of the random variable is $D_X = [0, 100]$. This is an uncountable subset of the real line. Thus, it is a continuous random variable.

**Probability**: Let $\Omega$ be the sample space and $\mathcal{U}$ be the class of all possible subsets of $\Omega$. As we defined earlier, a probability measure $P$, satisfies the following three conditions:

**a)** $P(\Omega) = 1$ **b)** $P(A) \geq 1$ for all $A \in \mathcal{U}$ **c)** $A_i \in \mathcal{U}$ and $A_i$'s are disjoint $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ .

Now, consider the previous example given in part (a) (tossing a coin three times). The sample space is $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. For any subset $A$ of $\Omega$, define a probability measure $P$ as $P(A) = n(A)/8$. Then we can define the following events:

$$A = \{observe\, no\, tail\} = \{HHH\} = \{w : X(w) = 0\} := \{X = 0\}$$

$$B = \{observe\, one\, tail\} = \{HHT, HTH, THH\} = \{w : X(w) = 1\} := \{X = 1\}$$

$$C = \{observe\, two\, tails\} = \{HTT, THT, TTH\} = \{w : X(w) = 2\} := \{X = 2\}$$

$$D = \{observe\, three\, tails\} = \{TTT\} = \{w : X(w) = 3\} := \{X = 3\}.$$

We can calculate the following probabilities,

$$P(A) = P(\{observe\, no\, tail\}) = P(\{HHH\}) = P(\{w : X(w) = 0\}) = P(\{X = 0\}) = 1/8$$

$$P(B) = P(\{observe\, one\, tail\}) = P(\{HHT, HTH, THH\}) = P(\{w : X(w) = 1\}) = P(\{X = 1\}) = 3/8$$

$$P(C) = P(\{observe\, two\, tails\}) = P(\{HTT, THT, TTH\}) = P(\{w : X(w) = 2\}) = P(\{X = 2\}) = 3/8$$

$$P(D) = P(\{observe\, three\, tails\}) = P(\{TTT\}) = P(\{w : X(w) = 3\}) = P(\{X = 3\}) = 1/8.$$

These probabilities are listed as a table below:

| $X = x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(X = x)$ | $1/8$ | $3/8$ | $3/8$ | $1/8$ |

We can also put these probabilities in a formula

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0,1,2,\ldots,n$$

where

$$p = 1/2 \ , \ n = 3 \ \text{ and } \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Now, we are ready to define the probability function of a random variable. Let $X$ be a discrete random variable with the range $D_X$ , then the probability function of a discrete random variable $X$ is given by

$$f(x) = \begin{cases} P(X = x) & , \quad x \in D_X \\ 0 & , \quad elsewhere. \end{cases}$$

Any function $f$ is a probability function of a random variable $X$ if it satisfies the following two conditions:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\sum_{x \in D_X} f(x) = 1$.

**Example**: Let $X$ be a random variable with the following probability function:

$$f(x) = \begin{cases} c\,x & , \quad x = 1,2,3,4,4 \\ 0 & , \quad elsewhere. \end{cases}$$

Since it is a probability function, we have

$$1 = \sum_{x=1}^{5} f(x) = \sum_{x=1}^{5} c\,x = c\sum_{x=1}^{5} x = c(1+2+3+4+5) = 15c$$

which implies that $c = 1/15$ and therefore the probability function can be written as

$$f(x) = \begin{cases} x/15 & , \quad x = 1,2,3,4,4 \\ 0 & , \quad elsewhere. \end{cases}$$

and we can calculate the probabilities as:

$$P(X \geq 4) = P(X = 4) + P(X = 5) = \frac{4}{15} + \frac{5}{15} = \frac{9}{15} = \frac{3}{5}$$

or

$$P(1 < X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{2}{15} + \frac{3}{15} + \frac{4}{15} = \frac{9}{15} = \frac{3}{5}.$$

**Cumulative Distribution Function**:

Let $X$ be a discrete random variable with the range $D_X$ . Then the cumulative distribution function of the fandom variable $X$ is a function from $\mathbb{R}$ to $[0,1]$ defined by

$$F:\mathbb{R}\to[0,1]$$
$$x\to F(x)=P(X\le x).$$

That is the cumulative distribution function can also be given as

$$F(x)=\sum_{t\le x} f(t)\ .$$

**Example**: Consider the coin tossing example again. When we toss a fair coin three times, the probability function was given above as:

| $X=x$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|
| $P(X=x)$ | $1/8$ | $3/8$ | $3/8$ | $1/8$ |

Note that all probabilities are greater than or equal to zero ( $f(x)\ge 0$ for all $x\in\mathbb{R}$ ) and the sum of the probabilities is $1$,

$$\sum_{x\in D_X} f(x)=P(X=0)+P(X=1)+P(X=2)+P(X=3)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1\ .$$

Now, let us find the cumulative distribution of the random variable $X$ (which counts the number of tails when we toss a fair coin three times). Note that,

if $x<0$ then $F(x)=P(X\le x)=P(\varnothing)=0$ ,

if $0\le x<1$ then $F(x)=P(X\le x)=P(X=0)=\dfrac{1}{8}$

if $1\le x<2$ then $F(x)=P(X\le x)=P(X=0)+P(X=1)=\dfrac{1}{8}+\dfrac{3}{8}=\dfrac{4}{8}$

if $2\le x<3$ then $F(x)=P(X\le x)=P(X=0)+P(X=1)+P(X=2)=\dfrac{1}{8}+\dfrac{3}{8}+\dfrac{3}{8}=\dfrac{7}{8}$

if $x\ge 3$ then $F(x)=P(X\le x)=P(X=0)+P(X=1)+P(X=2)+P(X=3)=\dfrac{1}{8}+\dfrac{3}{8}+\dfrac{3}{8}+\dfrac{1}{8}=1$ .

That is the cumulative distribution function of the random variable which is defined above is given below:

$$F(x)=\begin{cases} 0 & , \quad x<0 \\ 1/8 & , \quad 0\le x<1 \\ 4/8 & , \quad 1\le x<2 \\ 7/8 & , \quad 2\le x<3 \\ 1 & , \quad x\ge 0 \end{cases}$$

Figure 3.2

The probabilities can also be calculated from the cumulative distribution function as

$f(x) = P(X = x) = F(x^+) - F(x^-)$ for all $x \in \mathbb{R}$

where $F(x^+)$ is the right limit of the function $F$ at the point $x$ and $F(x^-)$ is the left limit of the function $F$ at the point $x$. Notice that, if $F$ is a the cumulative distribution of a discrete random variable, then it satisfies the following properties:

- $F$ is non-decreasing function (if $x_1 \le x_2$ then $F(x_1) \le F(x_2)$ )
- $F$ is right continuous , $\lim\limits_{h \to 0^+} F(x+h) = F(x)$
- $\lim\limits_{x \to \infty} F(x) = 1$ and $\lim\limits_{x \to -\infty} F(x) = 0$ .

**Continuous case**:

Let $X$ be a continuous random variable with the range $D_X$ . The function $f(x)$ is the probability density function of the random variable $X$ if

- $f(x) \ge 0$ for all $x \in \mathbb{R}$
- $\int\limits_{x \in D_X} f(x)dx = 1$

and the cumulative distribution function of the random variable (continuous) is given by

$F(x) = \int\limits_{-\infty}^{x} f(t)\,dt$ .

The cumulative distribution function of a continuous random variable $X$ satisfies the same properties given for the discrete case. Moreover, If $X$ is a continuous random variable then the cumulative distribution function is now continuous. For discrete case it was a right continuous function.

**Example**: a) Let $X$ be a random variable with the probability density function

$$f(x) = \begin{cases} c\,x & , \quad 0 < x < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Since it is a probability density function, we have

$$1 = \int\limits_{x \in D_X} f(x)dx = \int\limits_{0}^{1} f(x)\,dx = c\int\limits_{0}^{1} x\,dx = c\left.\frac{x^2}{2}\right|_{x=0}^{1} = \frac{c}{2}$$

and $c = 2$ . That is the probability density function of $X$ is

$$f(x) = \begin{cases} 2\,x & , \quad 0 < x < 1 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Using the probability density function we can calculate

$$P(0 < X < 0.5) = \int\limits_{0}^{0.5} 2x\,dx = x^2\Big|_{x=0}^{1/2} = \frac{1}{4}$$

$$P(0 < X < 1/3) = \int_0^{1/3} 2x\,dx = x^2\Big|_{x=0}^{1/3} = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$$

| | |
|---|---|
| In general, we can calculate any probability as<br>$P(X \in A) = \int_A f(x)\,dx$ .<br>This probability is the area of $A$ under the function (curve) $f(x)$ . | |
| | Figure 3.3 |

b) Now, let $X$ be a continuous random variable the the following probability density function:

$$f(x) = \begin{cases} c\,x^2 & , \quad 0 < x < 3 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Since it is a probability density function, we have

$$1 = \int_{x \in D_X} f(x)dx = \int_0^3 c\,x^2\,dx = c\frac{x^3}{3}\Big|_{x=0}^{3} = 9c$$

and $c = 1/9$ . That is the probability density function of $X$ is

$$f(x) = \begin{cases} x^2/9 & , \quad 0 < x < 3 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

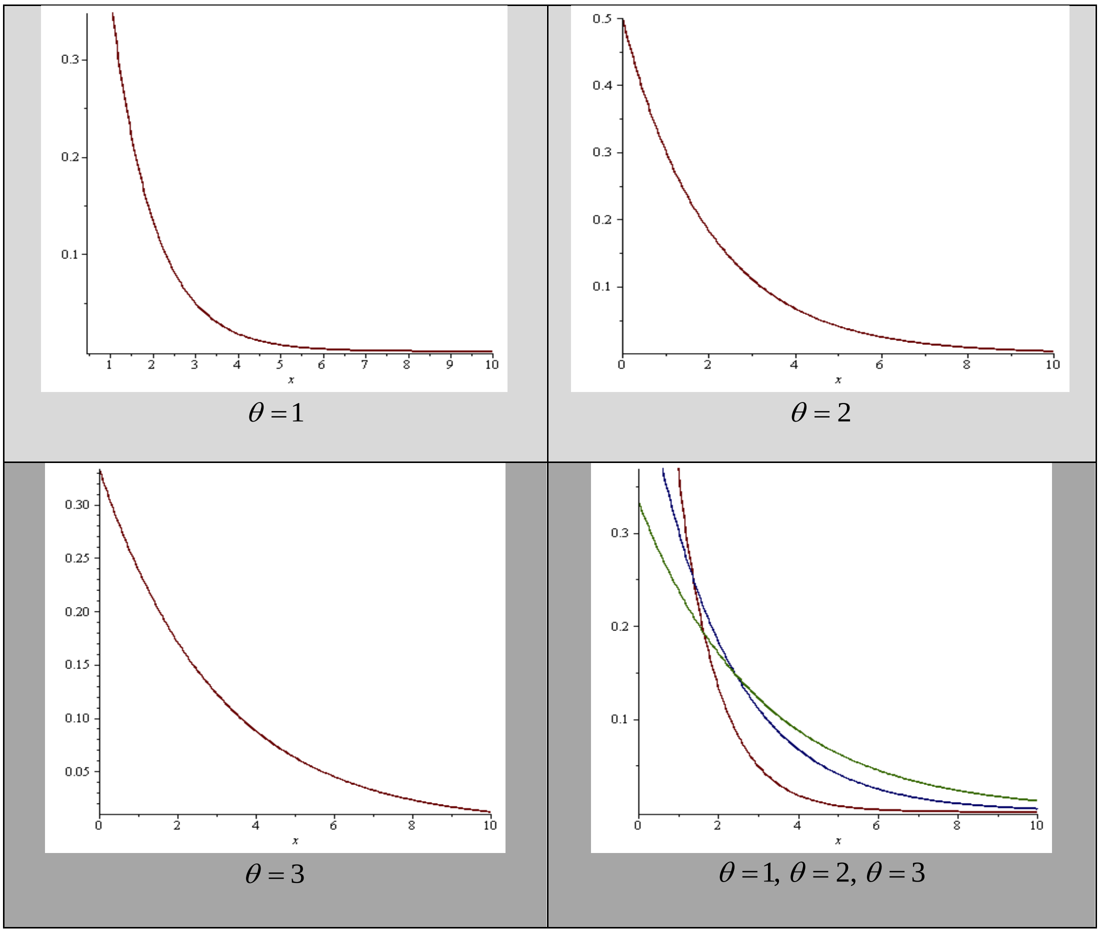
Using this probability density function (which is the area under the curve)

| | |
|---|---|
| $P(0 < X < 2) = \int_0^2 \frac{x^2}{9}dx = \frac{x^3}{27}\Big\|_{x=0}^{2} = \frac{8}{27}$<br>or similarly<br>$P(0 < X < 1) = \int_0^1 \frac{x^2}{9}dx = \frac{x^3}{27}\Big\|_{x=0}^{1} = \frac{1}{27}$ | |
| | |

Consider an experiment of measuring the life time of an electronic item. Let the life time (say in year) of an electronic item have a probability density function

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta} & , \quad x > 0 \\ 0 & , \quad elsewhere. \end{cases}$$

As we know that parameter (or parameters) $\theta$ characterize the distribution. Later, we are going to look at the ways of estimating and testing the parameter/parameters. Some of the graphs of the distribution (for different values of $\theta$ ) are given below.



$\theta = 1$ $\quad$ $\theta = 2$ $\quad$ $\theta = 3$ $\quad$ $\theta = 1,\ \theta = 2,\ \theta = 3$

Notice that, the shape of the probability function chances according to the values of $\theta$. Suppose you produce an electronic item and you want to put a warranty on it. The expected life time of the electronic item will be the warranty. Suppose you put 10-year warranty on it. If the expected life time is less than 10 years (say 5 years) you may sell more items but after 5 years they will be on the warranty and you have to replace all the items which last less than 10 years. That is, even you sell more items, you lose more money. Therefore, your estimation of your product plays an important role. You may estimate the expected life time of your product in many ways and therefore you need to estimate the expected life time as good as possible.

**Multivariate Case:**

In the above discussion we looked at a single variable. For example, if we want to look at the income variable we go outside and ask people's income to make an inference about the average income. However, when we go outside and ask people's income we can also ask their consumptions. The consumption is also another random variable. Therefore, on the same sample space (in our case people), we measure two different values (income and consumption). After we collect the data, we can make statistical inferences about the income and consumption. Moreover, we can also make statistical inferences about the relationship between income and consumption.

Let $\Omega$ be the sample space and we define a bivariate random vector (having components $X$ and $Y$) as

$$\begin{aligned}&(X,Y)':\Omega\to\mathbb{R}^2\\ &\qquad w\to(X,Y)'(w)=(X(w),Y(w))'.\end{aligned}$$

As we know, a random variable could be either discrete or continuous. For a bivariate case,

- Both random variables could be discrete
- Both random variables could be continuous
- One continuous and the other could be discrete.

In a similar way as we discussed for univariate case, we can define the joint probability function or joint probability density function for a bivariate random vector $(X,Y)'$. Let $X$ and $Y$ be two random variables having the ranges $D_X$ and $D_Y$ respectively and the range of the both random variables (joint) $D_{XY}$. Now there are three different cases:

- If both random variables are discrete, the joint probability function of $X$ and $Y$ is given by

  $$f(x,y)=\begin{cases}P(X=x,Y=y) & , \quad (x,y)\in D_{XY}\\ 0 & , \quad elsewhere.\end{cases}$$

  The joint probability function of $X$ and $Y$ is the one that satisfies

  a) $f(x,y)\geq 0$ for all $(x,y)\in\mathbb{R}^2$ b) $\sum\limits_{x\in D_X}\sum\limits_{y\in D_y} f(x,y)=1$

- If both random variables are continuous, the joint probability density function of $X$ and $Y$ ( $f(x,y)$ ) is the one that satisfies

  a) $f(x,y)\geq 0$ for all $(x,y)\in\mathbb{R}^2$ b) $\int\limits_{x\in D_X}\int\limits_{y\in D_Y} f(x,y)\,dy\,dx=1$

- If one is discrete (say $X$ ) and the other continuous, the joint probability function ( $f(x,y)$ ) is the one that satisfies

  a) $f(x,y)\geq 0$ for all $(x,y)\in\mathbb{R}^2$ b) $\sum\limits_{x\in D_X}\int\limits_{y\in D_Y} f(x,y)\,dy=1.$

**Example**: a) Discrete case: Let $X$ and $Y$ be two random variables with the joint probability function

$$f(x,y)=\begin{cases} c(x+y) & , \quad x=0,1,2\ ;\ y=1,2,3 \\ 0 & , \quad elsewhere. \end{cases}$$

The constant $c$ can be found as

$$1=\sum_{x=0}^{2}\sum_{y=1}^{3} c(x+y)=c\sum_{x=0}^{2}[(x+1)+(x+2)+(x+3)]=c\sum_{x=0}^{2}(3x+6)$$
$$=3c[(0+2)+(1+2)+(2+2)]=27c$$

which implies that $c=1/27$. That is, the joint probability function of $X$ and $Y$ is

$$f(x,y)=\begin{cases} (x+y)/27 & , \quad x=0,1,2\ ;\ y=1,2,3 \\ 0 & , \quad elsewhere. \end{cases}$$

Now, we can calculate probabilities as

$$P(X\le 1, Y=2)=P(X=0,Y=2)+P(X=1,Y=2)=[(0+2)+(1+2)]/27=5/27$$

or

$$P(X\le 1, Y\le 2)=P(X=0,Y=1)+P(X=0,Y=2)+P(X=1,Y=1)+P(X=1,Y=2)$$
$$=[(0+1)+(0+2)+(1+1)+(1+2)]/27=8/27.$$

b) Continuous case: Let $X$ and $Y$ be two random variables with the following joint probability density function:

$$f(x,y)=\begin{cases} c(x+y) & , \quad 0<x<2, 1<y<3 \\ 0 & , \quad elsewhere. \end{cases}$$

The constant $c$ is determined by the following integration:

$$1=\int_{x=0}^{2}\int_{y=1}^{3} c(x+y)\,dy\,dx=c\int_{x=0}^{2}\left(xy+y^2/2\right)_{y=1}^{3}dx=c\int_{x=0}^{2}[(3x+9/2)-(x+1/2)]dx$$
$$=c\int_{x=0}^{2}(2x+4)dx=c[x^2+4x]_{x=0}^{2}=12c.$$

This implies that $c=1/12$ and therefore the joint probability density function of $X$ and $Y$ can be written as:

$$f(x,y)=\begin{cases} (x+y)/12 & , \quad 0<x<2, 1<y<3 \\ 0 & , \quad elsewhere. \end{cases}$$

Using this joint probability density function, we can calculate $P(X\le 1, Y\le 2)$ as

$$P(X\le 1, Y\le 2)=\frac{1}{12}\int_{x=0}^{1}\int_{y=1}^{2}(x+y)\,dy\,dx=\frac{1}{12}\int_{x=0}^{1}\left(xy+y^2/2\right)_{y=1}^{2}dx=\frac{1}{12}\int_{x=0}^{1}[(x+3/2)]dx$$
$$=\frac{1}{12}\left[\frac{x^2}{2}+\frac{3x}{2}\right]_{x=0}^{1}=\frac{1}{12}\left[\frac{1}{2}+\frac{3}{2}\right]=\frac{1}{6}.$$

That is, $P(X\le 1, Y\le 2)=1/6$.

**Marginal, Independency and Conditional Probability Functions**:

Let $X$ and $Y$ be two random variables with the joint probability function (or joint probability density function) $f(x,y)$. If both random variables are discrete, then marginal probability function of $X$ and $Y$ can be found by,

$$f_X(x) = \sum_{y \in D_Y} f(x,y) \quad \text{for all } x \in \mathbb{R} \quad \text{and} \quad f_Y(y) = \sum_{x \in D_X} f(x,y) \quad \text{for all } y \in \mathbb{R}$$

and if both are continuous random variables then marginal probability density functions are calculated as

$$f_X(x) = \int_{y \in D_Y} f(x,y)\, dy \quad \text{for all } x \in \mathbb{R} \quad \text{and} \quad f_Y(y) = \int_{x \in D_X} f(x,y)\, dx \quad \text{for all } y \in \mathbb{R}.$$

**Definion**: a) Let $X$ and $Y$ be two random variables with the joint probability function (or joint probability density function) $f(x,y)$ with the marginal probability (or probability density function) $f_X(x)$ and $f_Y(y)$. If

$$f(x,y) = f_X(x) f_Y(y) \quad \text{for all } x, y \in \mathbb{R}$$

then the random variables $X$ and $Y$ are said to be independent.

b) The conditional probability (or probability density) function of $Y$ given $X = x$ is given by

$$f_{Y|X=x}(y \mid x) = \frac{f(x,y)}{f_X(x)} \quad \text{for all } x \in \mathbb{R} \text{ such that } f_X(x) > 0.$$

Similarly, the conditional probability (or probability density) function of $X$ given $Y = y$ is given by

$$f_{X|Y=y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \quad \text{for all } y \in \mathbb{R} \text{ such that } f_Y(y) > 0.$$

**Example**: Consider the previous example given above. That is, let $X$ and $Y$ be two random variables with the joint probability density function given by

$$f(x,y) = \begin{cases} (x+y)/12 & , \quad 0 < x < 2, 1 < y < 3 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

The marginal probability density functions are given below:

$$f_X(x) = \int_{y \in D_Y} f(x,y)\, dy = \frac{1}{12} \int_{y=1}^{3} (x+y) dy = \frac{1}{12} \left( xy + \frac{y^2}{2} \right)_{y=1}^{3} = \frac{1}{12}\big((3x + 9/2) - (x + 1/2)\big) = \frac{x+2}{6}$$

and

$$f_Y(y) = \int_{x \in D_X} f(x,y)\, dx = \frac{1}{12} \int_{x=0}^{2} (x+y) dy = \frac{1}{12} \left( \frac{x^2}{2} + xy \right)_{x=0}^{2} = \frac{(2+2y)}{12} = \frac{y+1}{6}.$$

Thus, the marginal probability density functions can be written as

$$f_X(x) = \begin{cases} (x+2)/6 & , \quad 0 < x < 2 \\ 0 & , \quad \textit{elsewhere} \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} (y+1)/6 & , \quad 1 < y < 3 \\ 0 & , \quad \textit{elsewhere.} \end{cases}$$

Since,

$$f(x,y)=\frac{x+y}{12}\neq\left(\frac{x+2}{6}\right)\left(\frac{y+1}{6}\right)=f_X(x)\ f_Y(y)$$

the random variables $X$ and $Y$ are NOT independent.

Now, let us try to find the conditional probability density function of $Y$ given $X=x$. Note that

$$f_{Y|X=x}(y\,|\,x)=\frac{f(x,y)}{f_X(x)}=\frac{(x+y)/12}{(x+2)/6}=\frac{x+y}{2(x+2)}$$

and therefore the conditional probability density function of $Y$ given $X=x$ is

$$f_{Y|X=x}(y\,|\,x)=\begin{cases}\dfrac{x+y}{2(x+2)} & ,\quad 1<y<3\\ 0 & ,\quad \textit{elsewhere.}\end{cases}$$

Notice that,

$$\int_{y=1}^{3} f_{Y|X=x}(y\,|\,x)\,dy=\int_{y=1}^{3}\frac{x+y}{2(x+2)}dy=\frac{1}{2(x+2)}\int_{y=1}^{3}(x+y)dy=\frac{1}{2(x+2)}\Big[xy+y^2/2\Big]_{y=1}^{3}$$

$$=\frac{1}{2(x+2)}\big[(3x+9/2)-(x+1/2)\big]=\frac{2x+4}{2(x+2)}=\frac{2(x+2)}{2(x+2)}=1.$$

Thus, it is a probability density function.