Chapter 3: Joint Distributions

1 Joint Distributions for Multiple Random Variables

- Very often, we are interested in more than one random variables simultaneously.
- For example, an investigator might be interested in both the height (H) and the weight (W) of an individual from a certain population.
- Another investigator could be interested in both the hardness (H) and the tensile strength (T) of a piece of cold-drawn copper.

Definition 1

- *Let E be an experiment and S be a corresponding sample space.*
- *Let X and Y be two functions each assigning a real number to each $s \in S$.*
- *We call (X,Y) a* ***two-dimensional random vector****, or a* ***two-dimensional random variable****.*

Similarly to one-dimensional situation, we can denote the **range space** of (X,Y) by

$$R_{X,Y} = \left\{ (x,y) \middle| x = X(s), y = Y(s), s \in S \right\}.$$

The definition above can be extended to more than two random variables.

Definition 2
Let $X_1, X_2, \ldots, X_n$ be n functions each assigning a real number to ***every outcome*** *$s \in S$. We call $(X_1, X_2, \ldots, X_n)$ an* ***n-dimensional random variable*** *(or an* ***n-dimensional random vector****).*

We define the discrete and continuous two-dimensional RVs as follows.

DEFINITION 3

1 (X,Y) *is a* ***discrete*** *two-dimensional RV if the number of possible values of* $(X(s),Y(s))$ *are finite or countable.*

That is the possible values of $(X(s),Y(s))$ *may be represented by*

$$(x_i, y_j), i = 1,2,3,\ldots; j = 1,2,3,\ldots$$

2 (X,Y) *is a* ***continuous*** *two-dimensional RV if the possible values of* $(X(s),Y(s))$ *can assume any value in some region of the Euclidean space* $\mathbb{R}^2$.

REMARK:

we can view X and Y separately to judge whether (X,Y) is discrete or continuous.

forward and backwards are true

- If both X and Y are discrete RVs, then (X,Y) is a discrete RV.
- Likewise, if both X and Y are continuous random variables, then (X,Y) is a continuous RV.
- Clearly, there are other cases. For example, X is discrete, but Y is continuous. These are not our focus in this module.

EXAMPLE 4 ((DISCRETE RANDOM VECTOR))

- Consider a TV set to be serviced.
- Let

$$X = \{\text{age to the nearest year of the set}\};$$

$$Y = \{\text{\# of defective components in the set}\}.$$

- (X,Y) is a discrete 2-dimensional RV.
- $R_{X,Y} = \{(x,y) | x = 0,1,2,\ldots; y = 0,1,2,\ldots,n\}$, where n is the total number of components in the TV.
- $(X,Y) = (5,3)$ means that the TV is 5 years old and has 3 defective components.

L–example 3.1

- A fast food restaurant operates a **drive-up facility** and a **walk-up window**.

- On a day, Let

$$X = \text{the proportion of time that the } \textbf{drive-up facility} \text{ is in use;}$$
$$Y = \text{the proportion of time that the } \textbf{walk-up window} \text{ is in use.}$$

- Then $R_{X,Y} = \{(x,y) | 0 \le x, 0 \le y \le 1\}$.
- (X,Y) is a continuous 2-dimensional RV.

Joint Probability Function

- We introduce the probability function for the discrete and continuous RVs separately.
- For discrete random vector, similar to the one-dimensional case, we define its probability function by associate a number with each possible value of the RV.

DEFINITION 5 (JOINT PROBABILITY FUNCTION FOR DISCRETE RV)
Let (X,Y) be a 2-dimensional ***discrete*** *RV, the* ***joint probability (mass) function*** *is defined by*

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

for x, y being possible values of X and Y, or in the other words $(x,y) \in R_{X,Y}$.

the set {X=x} is a subset of S, in the sense: {X=x} = {s E S : X(s) = x}

The joint probability mass function has the following properties:

(1) $f_{X,Y}(x,y) \ge 0$ for any $(x,y) \in R_{X,Y}$.

(2) $f_{X,Y}(x,y) = 0$ for any $(x,y) \notin R_{X,Y}$.

(3) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$;

or equivalently $\sum\sum_{(x,y) \in R_{X,Y}} f(x,y) = 1$.

(4) Let A be any subset of $R_{X,Y}$, then

$$P((X,Y) \in A) = \sum\sum_{(x,y) \in A} f_{X,Y}(x,y).$$

EXAMPLE 6
Find the value of k such that $f(x,y) = kxy$ for $x = 1,2,3$ and $y = 1,2,3$ can serve as a joint probability function.

Solution: $R_{X,Y} = \{(x,y) | x = 1,2,3; y = 1,2,3\}$.

$$\begin{array}{lll} f(1,1) = k, & f(1,2) = 2k, & f(1,3) = 3k, \\ f(2,1) = 2k, & f(2,2) = 4k, & f(2,3) = 6k, \\ f(3,1) = 3k, & f(3,2) = 6k, & f(3,3) = 9k. \end{array}$$

Based on property (3), we have $1 = \sum_{x=1}^{3}\sum_{y=1}^{3} kxy = k\sum_{x=1}^{3} x \sum_{y=1}^{3} y$

$$\begin{aligned} 1 &= \sum\sum_{(x,y)\in R_{X,Y}} f(x,y) \\ &= 1k+2k+3k+2k+4k+6k+3k+6k+9k, \end{aligned}$$

which results in $k = 1/36$.

L–example 3.2

- A company has 2 production lines, A and B, which produce at most 5 and 3 machines respectively.
- Let

$$\begin{aligned} X &= \text{number of machines produced by line } A \\ Y &= \text{number of machines produced by line } B. \end{aligned}$$

- The joint probability function $f(x,y)$ for (X,Y) is given in the table, where each entry represents $f(x_i, y_j) = P(X = x_i, Y = y_j)$.
- What is the probability that in a day line A produces more machines than line B?

Table for the joint probability function $f(x,y)$

y	x						Row Total
	0	1	2	3	4	5	
0	0	0.01	0.02	0.05	0.06	0.08	0.22
1	0.01	0.03	0.04	0.05	0.05	0.07	0.25
2	0.02	0.03	0.05	0.06	0.06	0.07	0.29
3	0.02	0.04	0.03	0.04	0.06	0.05	0.24
Column Total	0.05	0.11	0.14	0.20	0.23	0.27	1

Consider the event

$$A = \{\text{line } A \text{ produces more machines than line } B\} = \{X > Y\}.$$

Then we have

$$\begin{aligned}
P(A) &= P(X > Y) \\
&= P\Big((X,Y) = (1,0) \text{ or } (X,Y) = (2,0) \text{ or } \\
&\quad (X,Y) = (2,1) \text{ or } \ldots \text{ or } (X,Y) = (5,3)\Big) \\
&= P\Big((X,Y) = (1,0)\Big) + \ldots + P\Big((X,Y) = (5,3)\Big) \\
&= f(1,0) + f(2,0) + \ldots + f(5,3) = 0.73.
\end{aligned}$$

L–example 3.3

- A company has 9 executives; 4 are married, 3 have never married, and 2 are divorced.
- Three executives are to be randomly selected for promotion.
- Among the selective executives, let

$$\begin{aligned}
X &= \{\text{number of married executives}\} \\
Y &= \{\text{number of never married executives}\}.
\end{aligned}$$

- Find the joint probability function of X and Y.

Solution: Note that the executives are selected randomly; so every possible selection of the executives are equally likely.

- The total number of ways to select 3 executives out of 9 is $\binom{9}{3}$.
- The possible values of x and y are constrained by $x, y = 0, 1, 2, 3$ and $1 \leq x + y \leq 3$. The number of ways to select x married and y never married is given by $\binom{4}{x}\binom{3}{y}\binom{2}{3-x-y}$.
- Therefore, the joint probability function of (X,Y) is given by

$$\begin{aligned}
f_{X,Y}(x,y) &= P(X = x, Y = y) \\
&= \frac{\binom{4}{x}\binom{3}{y}\binom{2}{3-x-y}}{\binom{9}{3}},
\end{aligned}$$

  for $x, y = 0, 1, 2, 3$ such that $1 \leq x + y \leq 3$ and $f_{X,Y}(x,y) = 0$ otherwise.
- This joint p.f. can be summarized as a table.

x	y 0	1	2	3	Row Total
0	0	$3/84$	$6/84$	$1/84$	$10/84$
1	$4/84$	$24/84$	$12/84$	0	$40/84$
2	$12/84$	$18/84$	0	0	$30/84$
3	$4/84$	0	0	0	$4/84$
Column Total	$20/84$	$45/84$	$18/84$	$1/84$	1

Definition 7 (Joint probability function for continuous RV)
*Let (X,Y) be a 2-dimensional **continuous** RV; its **joint probability (density) function** is a function $f_{X,Y}(x,y)$ such that*

$$P((X,Y)\in D) = \int\int_{(x,y)\in D} f_{X,Y}(x,y)dydx,$$

for any $D \subset \mathbb{R}^2$. More specifically,

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{X,Y}(x,y)dydx.$$

The joint probability density function has the following properties:

(1) $f_{X,Y}(x,y) \ge 0$, for any $(x,y) \in R_{X,Y}$.

(2) $f_{X,Y}(x,y) = 0$, for any $(x,y) \notin R_{X,Y}$.

(3) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$;

or equivalently $\int\int_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)dxdy = 1$.

Example 8
Find the value c such that $f(x,y)$ below can serve as a joint p.d.f. for a RV (X,Y):

$$f(x,y) = \begin{cases} cx(x+y), & 0 \le x \le 1; 1 \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Solution: In order for $f(x,y)$ to be a p.d.f., we need

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx = \int_0^1\int_1^2 cx(x+y)dydx = c\int_0^1 x\left(x+\frac{1}{2}y^2\Big|_1^2\right)dx \\ &= c\int_0^1 x(x+1.5)dx = c\left(\frac{1}{3}x^3 + 1.5\cdot\frac{1}{2}x^2\right)\Big|_0^1 = c\cdot\frac{13}{12}, \end{aligned}$$

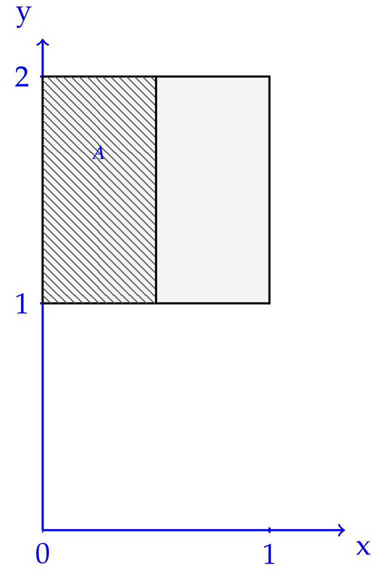
which implies $c = 12/13$.

L–example 3.4

Reuse the p.d.f. of Example 8:

$$f(x,y) = \begin{cases} \frac{12}{13}x(x+y), & 0 \le x \le 1; 1 \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}.$$

Assume that it is the joint p.d.f. of (X,Y). Let $A = \{(x,y)|0 < x < 1/2; 1 < y < 2\}$. Compute $P((X,Y) \in A)$.



- Set A corresponds to the shaped area in the figure on the right.
- We have

$$\begin{aligned} P((X,Y) \in A) &= P(0 < X < 1/2; 1 < Y < 2) \\ &= \int_0^{1/2} \int_1^2 \frac{12}{13}x(x+y)dydx \\ &= \frac{12}{13} \int_0^{1/2} x(x+1.5)dx \\ &= \frac{12}{13}\left(\frac{1}{3}x^3 + 1.5 \cdot \frac{1}{2}x^2\right)\Big|_0^{1/2} \\ &= 11/52. \end{aligned}$$

2 Marginal and Conditional Distributions

Definition 1 (Marginal Probability Distribution)
Let (X,Y) be a two-dimensional RV with joint p.f. $f_{X,Y}(x,y)$. We define the marginal distribution for X as follows.

- *If Y is a discrete RV, then for any x,*

$$f_X(x) = \sum_y f_{X,Y}(x,y).$$

- *If Y is a continuous RV, then for any x,*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

Remark:

- $f_Y(y)$ for Y is defined in the same way as that of X.

- We can view the marginal distribution as the "projection" of the 2D function $f_{X,Y}(x,y)$ to the 1D function.

- More intuitively, it is the distribution of X by ignoring the presence of Y.

  For example, consider a person of a certain community,

  - suppose $X =$ body weight, $Y =$ height. (X,Y) has a joint distribution $f_{X,Y}(x,y)$.
  - the marginal distribution $f_X(x)$ of X is the **distribution of body weights for all people in the community**.

- $f_X(x)$ should not involve the variable y; this can be viewed from its definition: y is either summed out or integrated over.

- $f_X(x)$ is a **probability function** so it satisfies all the properties of the probability function.

Example 2

- Revisit Example 6. The joint p.f. is given by $f(x,y) = \frac{1}{36}xy$ for $x = 1,2,3$ and $y = 1,2,3$.
- Note that X has three possible values: 1, 2, and 3. The marginal distribution for X is given by
  - for $x = 1$, $f_X(1) = f(1,1) + f(1,2) + f(1,3) = 6/36 = 1/6$.
  - for $x = 2$, $f_X(2) = f(2,1) + f(2,2) + f(2,3) = 12/36 = 1/3$.
  - for $x = 3$, $f_X(3) = f(3,1) + f(3,2) + f(3,3) = 18/36 = 1/2$.
  - for other values of x, $f_X(x) = 0$.
- Alternatively, for each $x \in \{1,2,3\}$,

$$
\begin{aligned}
f_X(x) &= \sum_y f(x,y) = \sum_{y=1}^{3} \frac{1}{36}xy \\
&= \frac{1}{36}x\sum_{y=1}^{3} y = \frac{1}{6}x.
\end{aligned}
$$

L–example 3.5

We reuse the joint p.f. of (X,Y) derived in L–Example 1:

x	y = 0	1	2	3	Row Total
0	0	$3/84$	$6/84$	$1/84$	$10/84$
1	$4/84$	$24/84$	$12/84$	0	$40/84$
2	$12/84$	$18/84$	0	0	$30/84$
3	$4/84$	0	0	0	$4/84$
Column Total	$20/84$	$45/84$	$18/84$	$1/84$	1

Can we read out the marginal p.f. of X and Y from the table directly?

L–example 3.6

Reuse the p.d.f. of Example 8:

$$f(x,y) = \begin{cases} \frac{12}{13}x(x+y), & 0 \le x \le 1; 1 \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}.$$

Assume that it is the joint p.d.f. of (X,Y). Find the marginal distribution of X.

<u>Solution</u>: (X,Y) is a continuous RV. For each $x \in [0,1]$, we have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_1^2 \frac{12}{13}x(x+y)dy \\ &= \frac{12}{13}x\left(x + \int_1^2 ydy\right) \\ &= \frac{12}{13}x(x+1.5); \end{aligned}$$

and for $x \notin [0,1]$, $f_X(x) = 0$.

Definition 3 (Conditional Distribution)

Let (X,Y) be a RV with joint p.f. $f_{X,Y}(x,y)$. Let $f_X(x)$ be the marginal p.f. for X. Then for any x such that $f_X(x) > 0$, the ***conditional probability function of*** Y ***given*** $X = x$ *is defined to be*

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

Remark:

- For any y such that $f_Y(y) > 0$, we can similarly define the **conditional distribution of** X **given** $Y = y$:

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

- $f_{Y|X}(y|x)$ is defined only for x such that $f_X(x) > 0$; likewise $f_{X|Y}(x|y)$ is defined only for y such that $f_Y(y) > 0$.

$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \geq 0$

- The practical meaning of $f_{Y|X}(y|x)$: the distribution of Y given that the random variable X is observed to take the value x.

$\sum_y f_{Y|X}(y|x) = \frac{\sum_y f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$

- Considering y as the variable (x as a fixed value), $f_{Y|X}(y|x)$ is a p.f., so it must satisfy all the properties of p.f..

- But $f_{Y|X}(y|x)$ is not a p.f. for x; this means that there is **NO** requirement $\int_{-\infty}^{\infty} f_{Y|X}(y|x)dx = 1$ for X continuous or $\sum_x f_{Y|X}(y|x) = 1$ for X discrete.

- With the definition, we immediately have
  - If $f_X(x) > 0$, $f_{X,Y}(x,y) = f_X(x)f_{Y|X}(y|x)$.
  - If $f_Y(y) > 0$, $f_{X,Y}(x,y) = f_Y(y)f_{X|Y}(x|y)$.

- One immediate application of the conditional distribution is to compute, for continuous RV,

$$\begin{aligned} P(Y \leq y | X = x) &= \int_{-\infty}^{y} f_{Y|X}(y|x)dy; \\ E(Y|X = x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy. \end{aligned}$$ ■

Their practical meanings are clear: the former is the probability that $Y \leq y$, given $X = x$; the latter is the average value of Y given $X = x$.

For discrete case, the computation is similarly established based on $f_{Y|X}(y|x)$; please fill in the details on your own.

EXAMPLE 4
Revisit Examples 6 and 2.

- The joint p.f. for (X,Y) is given by $f(x,y) = \frac{1}{36}xy$ for $x = 1,2,3$ and $y = 1,2,3$.
- The marginal p.f. for X is $f_X(x) = \frac{1}{6}x$ for $x = 1,2,3$.
- Therefore, $f_{Y|X}(y|x)$ is defined for any $x = 1,2$, or 3:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{(1/36)xy}{(1/6)x} = \frac{1}{6}y,$$

for $y = 1,2,3$.

- We can compute

$$\begin{aligned} P(Y=2|X=1) &= f_{Y|X}(2|1) = \frac{1}{6}\cdot 2 = 1/3; \\ P(Y\le 2|X=1) &= P(Y=1|X=1)+P(Y=2|X=1) \\ &= f_{Y|X}(1|1)+f_{Y|X}(2|1) = 1/6+1/3 = 1/2; \\ E(Y|X=2) &= 1\cdot f_{Y|X}(1|2)+2\cdot f_{Y|X}(2|2)+3\cdot f_{Y|X}(3|2) \\ &= 1\cdot(1/6)+2\cdot(2/6)+3\cdot(3/6) = 7/3. \end{aligned}$$

L–example 3.7

We reuse the joint p.f. of (X,Y) derived in L–Example 1:

x	y				Row
	0	1	2	3	Total
0	0	$3/84$	$6/84$	$1/84$	$10/84$
1	$4/84$	$24/84$	$12/84$	0	$40/84$
2	$12/84$	$18/84$	0	0	$30/84$
3	$4/84$	0	0	0	$4/84$
Column Total	$20/84$	$45/84$	$18/84$	$1/84$	1

Can we read out the conditional p.f. $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ from the table directly? How to compute $E(Y|X=x)$?

L–example 3.8 Reuse Examples 8 and L–Example 2.

- The joint p.f. for (X,Y) is given by

$$f(x,y)=\begin{cases}\frac{12}{13}x(x+y), & 0\le x\le 1; 1\le y\le 2\\ 0, & \text{elsewhere}\end{cases}.$$

- The marginal p.f. for X is given by

$$f_X(x)=\frac{12}{13}x(x+1.5),$$

  for $x\in[0,1]$.

- For each $x\in[0,1]$, the conditional p.f. $f_{Y|X}(y|x)$,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} = \frac{(12/13)x(x+y)}{(12/13)x(x+1.5)} \\ &= \frac{x+y}{x+1.5}, \end{aligned}$$

  for $y\in[1,2]$.

- We can compute

$$P(Y \leq 1.5|X = 0.5) = \int_1^{1.5} \frac{0.5+y}{0.5+1.5} dy = 0.5625.$$

- Furthermore

$$\begin{aligned} E(Y|X=0.5) &= \int_1^2 y \frac{0.5+y}{0.5+1.5} dy \\ &= \frac{1}{2} \int_1^2 (0.5y + y^2) dy \\ &= \frac{1}{2} \left(\frac{3}{4} + \frac{7}{3} \right) = 37/24. \end{aligned}$$

3 Independent Random Variables

Definition 1 (Independent Random Variables)

- *Random variables X and Y are **independent** if and only if for **any** x and y,*

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

- *Random variables $X_1, X_2, \ldots, X_n$ are **independent** if and only if for any $x_1, x_2, \ldots, x_n$,*

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

Remark:

- The above definition is applicable no matter whether (X,Y) is continuous or discrete.

- The "product feature" in the definition implies one necessary condition for independence: $R_{X,Y}$ needs to be a product space. In the sense that if X and Y are independent, for any $x \in R_X$ and any $y \in R_Y$, we have

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) > 0, \quad \blacksquare$$

  implying $R_{X,Y} = \{(x,y)|x \in R_X; y \in R_y\} = R_X \times R_Y$.

  <u>Conclusion</u>: if $R_{X,Y}$ is not a product space, then X and Y are not independent!

Properties of Independent Random Variables

Suppose X, Y are independent RVs.

(1) If A and B are arbitrary subsets of $\mathbb{R}$, the events $X \in A$ and $Y \in B$ are independent events in S. Thus

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B).$$

In particular, for any real numbers x, y,

$$P(X \le x; Y \le y) = P(X \le x)P(Y \le y).$$

(2) For arbitrary functions $g_1(\cdot)$ and $g_2(\cdot)$, $g_1(X)$ and $g_2(Y)$ are independent. For example,

- X^2 and Y are independent.
- $\sin(X)$ and $\cos(Y)$ are independent.
- e^X and $\log(Y)$ are independent.

(3) Independence is connected with conditional distribution.

- If $f_X(x) > 0$, then $f_{Y|X}(y|x) = f_Y(y)$.
- Likewise, if $f_Y(y) > 0$, then $f_{X|Y}(x|y) = f_X(x)$.

EXAMPLE 2

The joint p.f. of (X, Y) is given below.

x	y			$f_X(x)$
	1	3	5	
2	0.1	0.2	0.1	0.4
4	0.15	0.3	0.15	0.6
$f_Y(y)$	0.25	0.5	0.25	1

Are X and Y independent?

Solution:

- We need to check that for every x and y combination, whether we have

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

For example, from the table, we have $f_{X,Y}(2,1) = 0.1$; $f_X(2) = 0.4$, $f_Y(1) = 0.25$. Therefore

$$f_{X,Y}(2,1) = 0.1 = 0.4 \times 0.25 = f_X(2)f_Y(1).$$

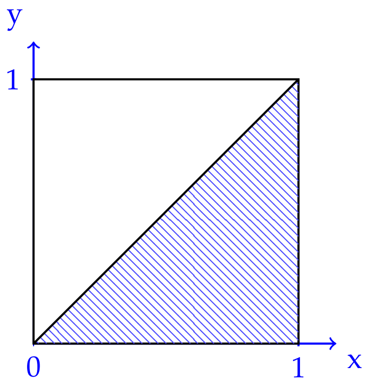
- In fact, we can check for each $x \in \{2,4\}$ and $y \in \{1,3,5\}$ combination, the equality holds.

- We conclude that X and Y are independent.

L–example 3.9 Given that

$$f_{X,Y}(x,y) = \begin{cases} 2(x+y), & \text{for } 0 \le x \le 1, 0 < y < x \\ 0 & \text{elsewhere} \end{cases}$$

Are X and Y independent?
Solution:



- The direct way of checking the independence is to check whether

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

holds for every (x,y) combination. The detail of this method is left as an exercise.

- For this question, we can immediately conclude that X and Y are not independent by checking that $R_{X,Y}$ is not a product space.

L–example 3.10 Suppose that (X,Y) is a discrete RV. The joint p.f. is given by

x	y				$f_X(x)$
	0	1	2	3	
0	1/8	1/4	1/8	0	1/2
1	0	1/8	1/4	1/8	1/2
$f_Y(y)$	1/8	3/8	3/8	1/8	1

Are X and Y independent?
Solution:
The zero entries in the table indicate that $R_{X,Y}$ is not a product space. Therefore, X and Y are not independent.

L–example 3.11 We have a handy way to check independence when $f_{X,Y}(x,y)$ has an explicit formula in $R_{X,Y}$.

X and Y are independent if and only if both of the following hold:

- $R_{X,Y}$, the range that the p.f. is positive, is a product space.

- For any $(x,y) \in R_{X,Y}$, we have $f_{X,Y}(x,y) = C \cdot g_1(x)g_2(y)$; that is, it can be "factorized" as the product of two functions g_1 and g_2, where the former **depends on** x **only**, the latter **depends on** y **only**, and C is a constant not depending on both x and y.

Note: $g_1(x)$ and $g_2(y)$ on their own are NOT necessarily p.f.s.

- We use the joint p.d. in Example 6 to illustrate: $f(x,y) = \frac{1}{36}xy$ for $x = 1,2,3$ and $y = 1,2,3$.
- $A_1 = \{1,2,3\}$ and $A_2 = \{1,2,3\}$, so the $R_{X,Y}$ is a product space.
- $f_{X,Y}(x,y) = \frac{1}{36} \cdot (x) \cdot (y)$: $C = 1/36$, $g_1(x) = x$, $g_2(y) = y$.
- We conclude that X and Y are independent.
- The advantage of this method is that we don't need to find the marginal distributions $f_X(x)$ and $f_Y(y)$ and check $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Following this strategy, we can get $f_X(x)$ and $f_Y(y)$ by standardizing $g_1(x)$ and $g_2(y)$. Consider $f_X(x)$ for illustration; $f_Y(y)$ is obtained similarly.

- If X is a discrete RV, its p.m.f. is given by

$$f_X(x) = \frac{g_1(x)}{\sum_{t \in R_X} g_1(t)}.$$

- If X is a continuous RV, its p.d.f. is given by

$$f_X(x) = \frac{g_1(x)}{\int_{t \in R_X} g_1(t)} dt.$$

- We continue to use the example above to illustrate. Here X is a discrete RV, $R_X = A_1 = \{1,2,3\}$. We obtain its p.m.f.:

$$f_X(x) = \frac{g_1(x)}{\sum_{x \in R_X} g_1(x)} = \frac{x}{\sum_{x=1}^{3} x} = x/6.$$

- Similarly, we get $f_Y(y) = y/6$.

L–example 3.12 Given that

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}x(1+y), & \text{for } 0 < x < 2, 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Are X and Y independent?
<u>Solution</u>:

- Set $A_1 = (0,2)$ and $A_2 = (0,1)$, then $R_{X,Y} = A_1 \times A_2$ is a product space.

- $f_{X,Y}(x,y)$ in $R_{X,Y}$ can be factorized by $C = 1/3$, $g_1(x) = x$, $g_2(y) = 1 + y$. Therefore, we conclude that X and Y are independent.

- Furthermore,

$$f_X(x) = \frac{g_1(x)}{\int_{x \in A_1} g_1(x)dx} = \frac{x}{\int_0^2 xdx} = x/2;$$

$$f_Y(y) = \frac{g_2(y)}{\int_{y \in A_2} g_2(y)dy} = \frac{1+y}{\int_0^1 (1+y)dy} = \frac{2}{3}(1+y).$$

4 Expectation and Covariance

Definition 1 (Expectation)
For any two variable function $g(x,y)$,

- *if (X,Y) is a discrete RV,*

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) f_{X,Y}(x,y);$$

- *if (X,Y) is a continuous RV,*

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dy dx.$$

If we let

$$g(X,Y) = (X - E(X))(Y - E(Y)) = (X - \mu_X)(Y - \mu_y),$$

the expectation $E[g(X,Y)]$ leads to the covariance of X and Y.

Definition 2 (Covariance)
*The **covariance** of X and Y is defined to be*

$$cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

$$= E[(X - \mu_x)(Y - \mu_y)]$$

- If X and Y are discrete RVs,

$$\text{cov}(X,Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f_{X,Y}(x,y).$$

- If X and Y are continuous RVs,

$$\text{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)dxdy.$$

The covariance has the following properties.

(1) $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$.

$= E[(x-\mu_x)(y-\mu_y)]$

can take out miu outside of the expecation since its a constant

$= E[xy - \mu_y \cdot x - \mu_x \cdot y + \mu_x \mu_y]$

$= E(xy) - \mu_y E(x) - \mu_x E(y) + \mu_x \mu_y$

$= Exy - \mu_y \mu_x - \mu_x \mu_y + \mu_x \mu_y$

(2) If X and Y are independent, then $\text{cov}(X,Y) = 0$. However, $\text{cov}(X,Y) = 0$ does not imply that X and Y are independent.

(3) $\text{cov}(aX+b, cY+d) = ac \cdot \text{cov}(X,Y)$.

cov(X,Y) = cov(Y,X)

cov(X+b, Y) = cov(X,Y)

cov(aX,Y) = a cov(X,Y)

(4) $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab \cdot \text{cov}(X,Y)$.

$V(ax) = a^2 V(X)$

$V(X+Y) = V(X) + V(Y) + 2 \text{ cov}(X, Y)$

EXAMPLE 3
Given the joint distribution for (X,Y):

x	$y=0$	$y=1$	$y=2$	$y=3$	$f_X(x)$
0	1/8	1/4	1/8	0	1/2
1	0	1/8	1/4	1/8	1/2
$f_Y(y)$	1/8	3/8	3/8	1/8	1

(a) Find $E(Y-X)$.

(b) Find $\text{cov}(X,Y)$.

<u>Solution</u>:

(a) Method 1:

$$\begin{aligned} E(Y-X) &= (0-0)(1/8) + (1-0)(1/4) + (2-0)(1/8) \\ &\quad + \ldots + (3-1)(1/8) = 1. \end{aligned}$$

Method 2:

$$E(Y-X) = E(Y) - E(X) = 1.5 - 0.5 = 1,$$

where

$$\begin{aligned} E(Y) &= 0\cdot(1/8) + 1\cdot(3/8) + 2\cdot(3/8) + 3\cdot(1/8) = 1.5 \\ E(X) &= 0\cdot(1/2) + 1\cdot(1/2) = 0.5. \end{aligned}$$

(b) We use $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$ to compute. Note that we have computed $E(X)$ and $E(Y)$ in Part (a).

$$\begin{aligned} E(XY) &= (0)(0)(1/8) + (0)(1)(1/4) + (0)(2)(1/8) \\ &\quad + \ldots + (1)(3)(1/8) = 1. \end{aligned}$$

Therefore

$$\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y) = 1 - (0.5)(1.5) = 0.25.$$

L–example 3.13 Suppose that (X,Y) has the p.f.

$$f_{X,Y}(x,y) = \begin{cases} x^2 + \dfrac{xy}{3}, & \text{for } 0 \le x \le 1, 0 \le y \le 2 \\ 0, & \text{otherwise} \end{cases}.$$

(a) Find $f_X(x)$, $f_Y(y)$ and $f_{Y|X}(y|x)$.

(b) Find $\mathrm{cov}(X,Y)$.

Solution:

(a) We first find the marginal density of X.

For $0 \le x \le 1$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = \int_0^2 \left(x^2 + \frac{xy}{3}\right)\mathrm{d}y \\ &= \left(x^2 y + \frac{xy^2}{6}\right)\Bigg|_{y=0}^{2} = 2x^2 + \frac{2x}{3}. \end{aligned}$$

It is clear that $f_X(x) = 0$ for $x < 0$ or $x > 1$. Thus

$$f_X(x) = \begin{cases} 2x^2 + \dfrac{2x}{3}, & \text{for } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}.$$

Similarly, the marginal density of Y is given as

$$f_Y(y) = \begin{cases} \frac{1}{3} + \frac{y}{6}, & \text{for } 0 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}.$$

The conditional probability density function of Y given $X = x$ when $0 \leq x \leq 1$ is then given as

$$\begin{aligned} f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} &= \begin{cases} \frac{x^2 + xy/3}{2x^2 + 2x/3}, & \text{for } 0 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{3x+y}{2(3x+1)}, & \text{for } 0 \leq y \leq 2 \\ 0, & \text{otherwise} \end{cases}. \end{aligned}$$

(b) We shall use the expression $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$.

Now

$$\begin{aligned} E(XY) &= \int_0^2 \int_0^1 xy\left(x^2 + \frac{xy}{3}\right) \text{dx dy} \\ &= \int_0^2 \int_0^1 \left(yx^3 + \frac{y^2x^2}{3}\right) \text{dx dy} \\ &= \int_0^2 \left(y\frac{x^4}{4} + \frac{y^2x^3}{9}\right)\Big|_{x=0}^{1} \text{dy} \\ &= \int_0^2 \left(\frac{y}{4} + \frac{y^2}{9}\right) \text{dy} \\ &= \frac{43}{54}. \end{aligned}$$

We have computed the marginal distributions for X and Y in Part (a). Thus

$$E(X) = \int_0^1 x\left(2x^2 + \frac{2x}{3}\right) \text{dx} = \left(\frac{2x^4}{4} + \frac{2x^3}{9}\right)\Big|_{x=0}^{1} = \frac{13}{18},$$

and

$$E(Y) = \int_0^2 y\left(\frac{1}{3} + \frac{y}{6}\right) \text{dy} = \left(\frac{y^2}{6} + \frac{y^3}{18}\right)\Big|_{y=0}^{2} = \frac{10}{9}.$$

This gives

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{43}{54} - \frac{13}{18} \times \frac{10}{9} = -\frac{1}{162}.$$

L–example 3.14

- Start from $V(X+Y) = V(X) + V(Y) + 2\operatorname{cov}(X,Y)$, we can have some interesting results.
- By induction, we have for any random variables $X_1, X_2, \ldots, X_n$,

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n) + 2\sum_{j>i} \operatorname{cov}(X_i, X_j).$$

- If X and Y are independent, we have

$$V(X \pm Y) = V(X) + V(Y).$$

- By induction, we have if $X_1, X_2, \ldots, X_n$ are independent,

$$V(X_1 \pm X_2 \pm \ldots \pm X_n) = V(X_1) + V(X_2) + \ldots + V(X_n).$$