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.

- 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 3.1 (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.

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 4 (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 joint probability mass function has the following properties:

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

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

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

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 3.2 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.
<u>Solution</u>: $R_{X,Y} = \{(x,y)|x = 1,2,3; y = 1,2,3\}$.

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

Based on property (3), we have

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

which results in $k = {}^{1}/_{36}$.

DEFINITION 5 (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) $f_{X,Y}(x,y) \geq 0$, for any $(x,y) \in R_{X,Y}$.

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

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

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

Example 3.3 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 $c = 12/13$.

2 Marginal and Conditional Distributions

Definition 6 (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 3.4

- Revisit Example 3.2. 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 $x \in \{1,2,3\}$,

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

DEFINITION 7 (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 $f_Y(y) > 0$.

- 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.

- 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,

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

  Their practical meanings are clear: the former is the probability that $Y \le y$, 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 3.5 Revisit Examples 3.2 and 3.4.

- The joint p.f. for (X,Y) is given by $f(x,y)=\dfrac{1}{36}xy$ for $x=1,2,3$ and $y=1,2,3$.
- The marginal p.f. for X is $f_X(x)=\dfrac{1}{6}x$ 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

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

3 Independent Random Variables

Definition 8 (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

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

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

Conclusion: 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,

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

(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 3.6 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.

DEFINITION 9 (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 10 (COVARIANCE)
The ***covariance*** *of X and Y is defined to be*

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

- 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,

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

The covariance has the following properties.

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

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

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

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

Example 3.7 Given the joint distribution for (X, Y):

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

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

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

Solution:

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

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