Three

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 individuals 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 (Two-dimensional Random Vector)

Let E be an experiment and S be a corresponding sample space. Suppose X and Y are 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****.*

DEFINITION 2 (RANGE SPACE)
Similar to the one-dimensional situation, we can denote the ***range space*** *of* (X,Y) *by*

$$R_{X,Y} = \{(x,y) \mid x = X(s), y = Y(s), s \in S\}.$$

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

Definition 3 (n-dimensional Random Vector)
Let $X_1, X_2, \ldots, X_n$ be n functions each assigning a real number to ***every outcome*** *$s \in S$.*

We call $(X_1, X_2, \ldots, X_n)$ a ***n-dimensional random vector****, or a* ***n-dimensional random variable****.*

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

DEFINITION 4

(X,Y) *is a* ***discrete two-dimensional random variable*** *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*

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

(X,Y) *is a* ***continuous two-dimensional random variable*** *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 random variables, then (X, Y) is discrete.
- If both X and Y are continuous random variables, then (X, Y) is continuous.
- Clearly, there are other cases. For example, X is discrete, but Y is continuous. These are not the focus of this course.

EXAMPLE 3.1
Consider a TV set that needs to be serviced.

Let X be the age of the set, rounded to the nearest year, and Y be the numbers of defective components in the set.

Then (X,Y) is a discrete 2-dimensional random variable and its range space is given as

$$R_{X,Y} = \{(x,y) \mid x = 0,1,2,\ldots; y = 0,1,2,\ldots,n\},$$

where n is the total number of components in the TV.

For example, $(X,Y) = (5,3)$ means that the TV is 5 years old and has 3 defective components.

Joint Probability Function

We will now introduce the probability functions for discrete and continuous random vectors.

For the discrete random vector, similar to the one-dimensional case, we define its probability function by associating a number with each possible value of the random variable.

Definition 5 (Discrete joint probability function)
Let (X, Y) be a 2-dimensional ***discrete*** *random variable. Its* ***joint probability (mass) function*** *is defined by*

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

for $(x,y) \in R_{X,Y}$.

PROPERTIES OF THE DISCRETE JOINT PROBABILITY FUNCTION

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

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

$$f(x,y) = kxy, \quad \text{for } x = 1,2,3 \text{ and } y = 1,2,3,$$

can serve as a joint probability function.

Solution:
Note that $R_{X,Y} = \{(x,y) \mid x = 1,2,3; y = 1,2,3\}$, and

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

Using Property (3), we have

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

This results in $k = 1/36$.

DEFINITION 6 (CONTINUOUS JOINT PROBABILITY FUNCTION)
Let (X,Y) be a 2-dimensional ***continuous*** *random variable. Its* ***joint probability (density) function*** *is a function $f_{X,Y}(x,y)$ such that*

$$P((X,Y) \in D) = \iint_{(x,y)\in D} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x,$$

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

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

Properties of the continuous joint probability function

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) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{dx}\,\mathrm{dy} = 1.$

Equivalently, $\displaystyle\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)\,\mathrm{dx}\,\mathrm{dy} = 1.$

EXAMPLE 3.3

Find the value c such that $f(x,y)$ below can serve as a joint probability density function for a random variable (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 probability density function, we need

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

This implies that $c = 12/13$.

2 Marginal and Conditional Distributions

We now consider the marginal distributions.

Put simply, the marginal distribution of X is the individual distribution of X, ignoring the value of Y.

DEFINITION 7 (MARGINAL PROBABILITY DISTRIBUTION)

Let (X, Y) be a two-dimensional random variable with joint probability function $f_{X,Y}(x, y)$. We define the ***marginal distribution*** *of X as follows.*

If Y is a discrete random variable, then for any x,

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

If Y is a continuous random variable, then for any x,

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

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

  For example, consider a person from a certain community.

  - Suppose $X =$ body weight, $Y =$ height, and (X,Y) has 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
We revisit Example 3.2. The joint probability function is given by

$$f(x,y) = \frac{1}{36}xy, \quad \text{for } x = 1,2,3 \text{ and } y = 1,2,3.$$

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 $x \in \{1, 2, 3\}$,

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

DEFINITION 8 (CONDITIONAL DISTRIBUTION)
Let (X, Y) be a random variable with joint probability function $f_{X,Y}(x, y)$. Let $f_X(x)$ be the marginal probability function 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 $Y = y$** as

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

- $f_{Y|X}(y|x)$ 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 intuitive 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 (and x as a fixed value), $f_{Y|X}(y|x)$ is a probability function, so it must satisfy all the properties of a probability function.

- However, $f_{Y|X}(y|x)$ is not a probability function for x. This means that there is **NO** requirement that

  - $\int_{-\infty}^{\infty} f_{Y|X}(y|x)\,\mathrm{dx} = 1$, for X continuous; or
  - $\sum_x f_{Y|X}(y|x) = 1$, for X discrete.

- With this 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 random variable,

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

  Their interpretations 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 the discrete case, the results can be similarly established, based on the definition of $f_{Y|X}(y|x)$.

EXAMPLE 3.5

We revisit Examples 3.2 and 3.4. The joint probability function for (X, Y) is given by

$$f(x, y) = {}^{xy}\!/_{36}, \quad \text{for } x = 1, 2, 3 \text{ and } y = 1, 2, 3.$$

The marginal probability function for X is

$$f_X(x) = {}^{x}\!/_{6}, \quad \text{for } x = 1, 2, 3.$$

Therefore $f_{Y|X}(y|x)$ is defined for any $x = 1, 2, 3$:

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{({}^{xy}\!/_{36})}{({}^{x}\!/_{6})} = {}^{y}\!/_{6}, \quad \text{for } y = 1, 2, 3.$$

We can also compute

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

3 Independent Random Variables

We next discuss independence for random variables.

Definition 9 (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 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 random variables.

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

EXAMPLE 3.6

The joint probability function 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. Therefore X and Y are independent.

4 Expectation and Covariance

Similar to one dimensional random variable, we can talk about the expectation of a random vector.

Definition 10 (Expectation)
Consider any two variable function $g(x,y)$.

If (X,Y) is a discrete random variable,

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

If (X,Y) is a continuous random variable,

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

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

$$\operatorname{cov}(X,Y) = E[(X - E(X))(Y - E(Y))].$$

Remark

If X and Y are discrete random variables,

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

If X and Y are continuous random variables,

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

PROPERTIES OF THE COVARIANCE

The covariance has the following properties.

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

This is true because

$$\begin{aligned}\operatorname{cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E[XY - Y\mu_X - X\mu_Y + \mu_X\mu_Y] \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y.\end{aligned}$$

(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.

Take note that the two statements can be summarised as:

(i) $X \perp Y \Rightarrow \operatorname{cov}(X,Y) = 0$;

(ii) $X \perp Y \not\Leftarrow \operatorname{cov}(X,Y) = 0$.

For (i), note that if X and Y are independent, then $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. So

$$
\begin{aligned}
E(XY) &= \sum_i \sum_j x_i y_j f_{X,Y}(x_i, y_j) = \sum_i \sum_j x_i y_j f_X(x_i) f_Y(y_j) \\
&= \sum_i x_i f_X(x_i) \sum_j y_j f_Y(y_j) = E(X)E(Y).
\end{aligned}
$$

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

This can be derived using the following 3 formulas:

(i) $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$;

(ii) $\operatorname{cov}(X+b,Y) = \operatorname{cov}(X,Y)$;

(iii) $\operatorname{cov}(aX,Y) = a\operatorname{cov}(X,Y)$.

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

This can be derived using the following 2 formulas:

(i) $V(aX) = a^2V(X)$;

(ii) $V(X+Y) = V(X) + V(Y) + 2\operatorname{cov}(X,Y)$.

EXAMPLE 3.7

We are 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

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

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

Solution:

(i) 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

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

(ii) 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 (i).

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

Therefore

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