

## 5 Joint Probability Distributions

### 5.1 Joint Probability Distributions for Two Random Variables

**Definition.** Let  $X, Y$  be two random variables. With a ordered pair  $(X, Y)$ , we define a **joint probability distribution function** and denote

$$f_{XY}(x, y).$$

In discrete case, we called **joint probability mass function**; In continuous case, we called **joint probability density function**.

**Property.** Let  $X, Y$  be two random variables.

1. If  $f_{XY}(x, y)$  is a joint probability mass function, then it satisfies the following properties:

- (1)  $f_{XY}(x, y) \geq 0$

- (2)  $\sum_{x \in X} \sum_{y \in Y} f_{XY}(x, y) = 1$

- (3)  $f_{XY}(x, y) = P(X = x, Y = y) = P(X = x \ \& \ Y = y)$

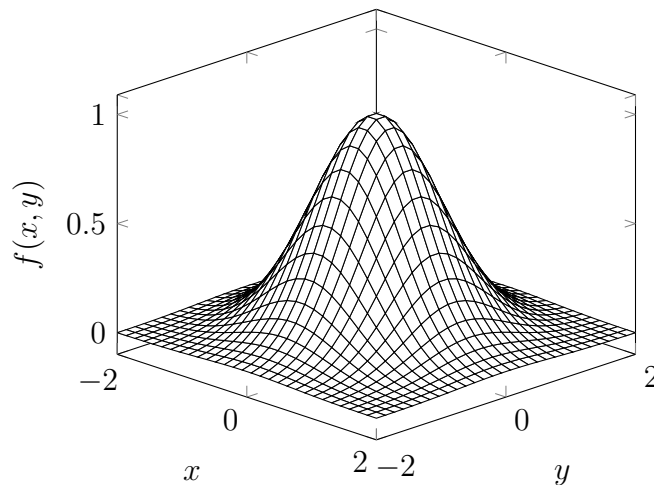
2. If  $f_{XY}(x, y)$  is a joint probability density function, then it satisfies the following properties:

- (1)  $f_{XY}(x, y) \geq 0$  for all  $x, y$

- (2)  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$

- (3) For any region  $R$  of two-dimensional space,

$$f_{XY}(x, y) = P((X, Y) \in R) = \iint_R f_{XY}(x, y) dx dy.$$



**Definition.** (Marginal Density) Let  $X, Y$  be two random variables.

1. For discrete case, the marginal(邊際) density is defined as

$$(1) f_X(x) = \sum_{y \in Y} f_{XY}(x, y)$$

$$(2) f_Y(y) = \sum_{x \in X} f_{XY}(x, y)$$

2. For continuous case, the marginal density is defined as

$$(1) f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

$$(2) f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

**Example.** Suppose  $f_{XY}(x, y)$  is a joint probability mass function with two random variables  $X = \{0, 1, 2\}$  and  $Y = \{0, 1, 2, 3\}$ . It is as shown in the table below

$Y = y$	$X = x$		
	0	1	2
0	0.15	0.1	0.05
1	0.02	0.1	0.05
2	0.02	0.03	0.2
3	0.01	0.02	0.25

Then the marginal probability distributions of  $X$  and  $Y$  are

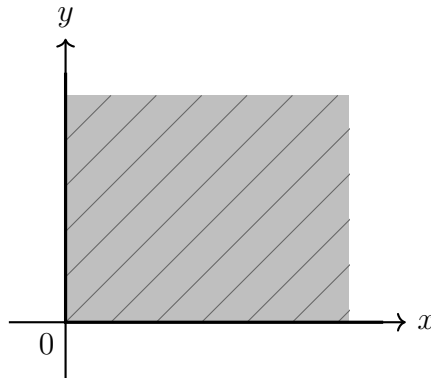
$Y = y$	$X = x$			$f_Y(y)$
	0	1	2	
0	0.15	0.1	0.05	<b>0.3</b>
1	0.02	0.1	0.05	<b>0.17</b>
2	0.02	0.03	0.2	<b>0.25</b>
3	0.01	0.02	0.25	<b>0.28</b>
$f_X(x)$	<b>0.2</b>	<b>0.25</b>	<b>0.55</b>	1

**Example.** Let  $f_{XY}(x, y)$  be a joint probability density function and be defined as

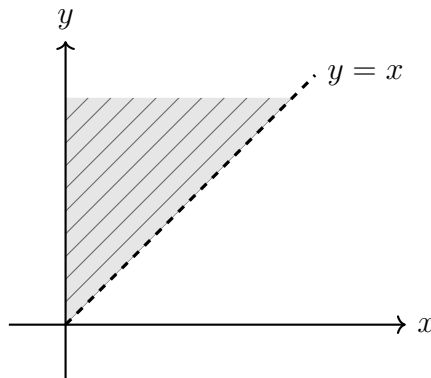
$$f_{XY}(x, y) = ke^{-0.01x-0.02y} \text{ for } 0 < x < y < \infty,$$

where  $k = 6 \times 10^{-6}$ . The region with nonzero probability is shaded in below. Find the  $P(Y > 2000)$ .

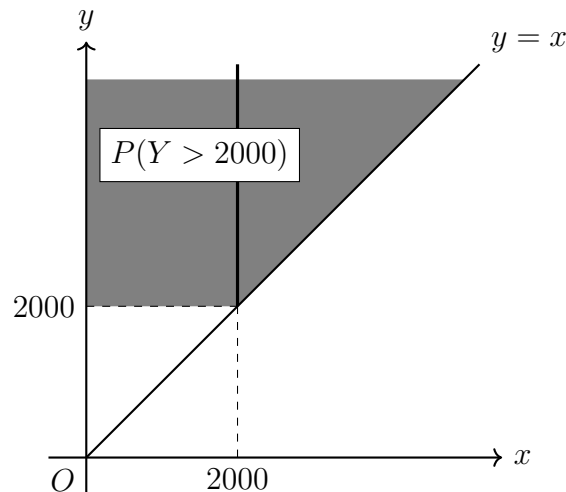
(i) **Region :**  $0 < x < \infty, 0 < y < \infty$



(ii) **Region :**  $0 < x < y < \infty$



**Sol.** The region is shaded in below.



So

$$\begin{aligned}
P(Y > 2000) &= \int_{2000}^{\infty} \int_0^y f_{XY}(x, y) dx dy \\
&= \int_{2000}^{\infty} \int_0^y k e^{-0.01x-0.02y} dx dy \\
&= k \int_{2000}^{\infty} e^{-0.02y} \left[ \int_0^y e^{-0.01x} dx \right] dy \\
&= k \int_{2000}^{\infty} e^{-0.02y} \left[ \frac{e^{-0.01x}}{-0.01} \right]_0^y dy \\
&= k \int_{2000}^{\infty} e^{-0.02y} \cdot 100(1 - e^{-0.01y}) dy \\
&= 100k \int_{2000}^{\infty} (e^{-0.02y} - e^{-0.03y}) dy \\
&= 100k \lim_{b \rightarrow \infty} \left[ \frac{e^{-0.02y}}{-0.02} - \frac{e^{-0.03y}}{-0.03} \right]_{2000}^b \\
&= 100k \left[ (0 - 0) - \left( \frac{e^{-0.02(2000)}}{-0.02} - \frac{e^{-0.03(2000)}}{-0.03} \right) \right] \\
&= 100k \left( 50e^{-40} - \frac{100}{3}e^{-60} \right)
\end{aligned}$$

Substitute  $k = 6 \times 10^{-6}$  :

$$\begin{aligned}
P(Y > 2000) &= 6 \times 10^{-4} \left( 50e^{-40} - \frac{100}{3}e^{-60} \right) \\
&= 0.03e^{-40} - 0.02e^{-60}.
\end{aligned}$$

## 5.2 Conditional Probability Distributions and Independence

### 5.2.1 Independence

**Definition.** Let  $X$  and  $Y$  be two random variables. Then they are **independent** if and only if

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

**Remark.** Recall math in high school, we know that

$$P(A \cap B) = P(A)P(B)$$

if both events  $A$  and  $B$  are independent.

**Definition.** Let  $X$  and  $Y$  be two random variables. Then the mean from a joint distribution  $(X, Y)$  with a transformer function  $H(X, Y)$  of  $X$  and  $Y$  is defined as

$$E[H(X, Y)] = \sum_{x \in X} \sum_{y \in Y} H(X, Y) f_{XY}(x, y) \quad (\text{Discrete case})$$

$$E[H(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(X, Y) f_{XY}(x, y) dx dy \quad (\text{Continuous case})$$

**Example.** Find  $\sum_{x \in X} \sum_{y \in Y} x f_{XY}(x, y)$ .

**Sol.**

$$\begin{aligned} \sum_{x \in X} \sum_{y \in Y} x f_{XY}(x, y) &= \sum_{x \in X} x \left( \sum_{y \in Y} f_{XY}(x, y) \right) \\ &= \sum_{x \in X} x f_X(x) \\ &= E(X). \end{aligned}$$

**Example.** Find  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dy dx$ .

**Sol.**

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dy dx &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_{XY}(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= E(X). \end{aligned}$$

### 5.2.2 Covariance and Correlation

Let's review the variance in one variable.

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E^2(X)) \\ &= E(X^2) - 2E^2(X) + E^2(X) \\ &= E(X^2) - E^2(X) \end{aligned}$$

So we have

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E^2(X) \\ &= E(X \cdot X) - E(X) \cdot E(X) \\ &\Rightarrow E(X \cdot Y) - E(X) \cdot E(Y). \end{aligned}$$

$E(X \cdot Y) - E(X) \cdot E(Y) = E[(X - E(X))(Y - E(Y))]$ , it is covariance(共變異數).

**Definition.** The **covariance** between the random variables  $X$  and  $Y$ , denoted as  $\text{cov}(X, Y)$  or  $\sigma_{XY}$  is

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

**Theorem.** Let  $(X, Y)$  be a ordered pair random variables with joint probability distribution function  $f_{XY}(x, y)$ . If  $X$  and  $Y$  are independent, then  $\text{cov}(X, Y) = 0$ .

**Definition.** (Correlation coefficient) The **Pearson correlation coefficient**(皮爾森相關係數) between random variables  $X$  and  $Y$ , denoted as  $\rho_{XY}$ , is

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

For any two random variables  $X$  and  $Y$ ,

$$-1 \leq \rho_{XY} \leq 1.$$

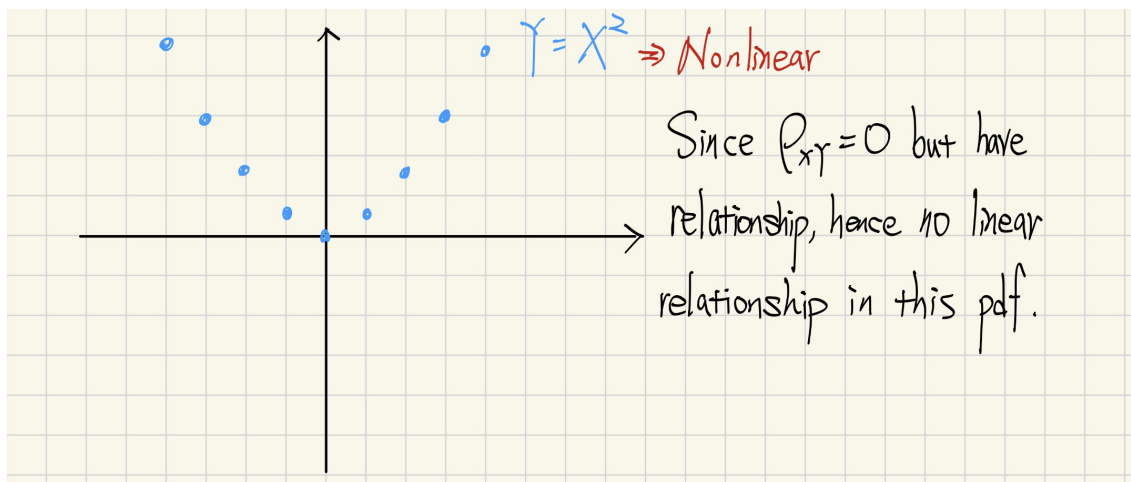
**Theorem.** The conditions for **linearity** between  $X$  and  $Y$  if and only if

$$Y = \beta_0 + \beta_1 X.$$

**Remark.** Let  $\rho_{XY}$  be the Pearson correlation coefficient between random variables  $X$  and  $Y$ .

1. If  $\rho_{XY} = 1$ , then we call it is perfect positive correlation and  $\beta_1 > 0$  for  $Y = \beta_0 + \beta_1 X$ .
2. If  $\rho_{XY} = -1$ , then we call it is perfect negative correlation and  $\beta_1 < 0$  for  $Y = \beta_0 + \beta_1 X$ .
3. If  $\rho_{XY} = 0$ , then we call it is no linear relationship.

**Example.** Consider  $Y = X^2$ .



**Remark.** If  $X$  and  $Y$  are independent, then  $\rho_{XY} = 0$ .

### 5.2.3 Conditional Probability Distributions

**Definition.** Let  $(X, Y)$  be a ordered pair random variables with joint probability distribution function  $f_{XY}(x, y)$ . Suppose  $f_X(x)$  and  $f_Y(y)$  are marginal probability distribution function for  $f_{XY}(x, y)$ .

(1) The conditional density for  $X$  given  $Y = y$  is defined as

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

Recall math in high school, we know that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

(2) The conditional density for  $Y$  given  $X = x$  is defined as

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

**Example.** Suppose

$$f_{XY}(x, y) = \frac{c}{x}, \quad c = \frac{1}{6 - 27 \ln(33/27)},$$

where  $27 \leq y \leq x \leq 33$ . Find

- (a)  $f_X(x)$
- (b)  $f_Y(y)$
- (c)  $f_{X|y}(x)$
- (d) The probability that  $x > 32$  given  $y = 30$ .
- (e) The expected value of  $X$  given  $y = 30$ .

**Sol.** We don't substitute  $c$  into the function since it is too difficult to evaluate.

(a)

$$\begin{aligned} f_X(x) &= \int_{27}^x f_{XY}(x, y) dy \\ &= \int_{27}^x \frac{c}{x} dy \\ &= \frac{cy}{x} \Big|_{27}^x \\ &= c \frac{x - 27}{x}, \text{ where } 27 \leq x \leq 33. \end{aligned}$$

(b)

$$\begin{aligned} f_Y(y) &= \int_y^{33} f_{XY}(x, y) dx \\ &= \int_y^{33} \frac{c}{x} dx \\ &= c(\ln 33 - \ln y) \\ &= c \ln \frac{33}{y}, \text{ where } 27 \leq y \leq 33. \end{aligned}$$

(c)

$$\begin{aligned} f_{X|y}(x) &= \frac{f_{XY}(x, y)}{f_Y(y)} \\ &= \frac{c/x}{c(\ln 33 - \ln y)} \\ &= \frac{1}{x(\ln 33 - \ln y)} \\ &= \frac{1}{x \ln(33/y)}, \text{ where } y \leq x \leq 33. \end{aligned}$$

(d)

$$\begin{aligned} P(X > 32|y = 30) &= \int_{32}^{33} \frac{1}{x(\ln 33/30)} dx \\ &= \left[ \frac{\ln |x|}{\ln(33/30)} \right]_{32}^{33} \\ &= \frac{\ln 33 - \ln 32}{\ln(33/30)} \\ &= \frac{\ln(33/32)}{\ln(33/30)}. \end{aligned}$$

(e)

$$\begin{aligned} E(X|y = 30) &= \int_{30}^{33} x f_{X|y=30}(x) dx \\ &= \int_{30}^{33} x \frac{1}{x \ln(33/30)} dx \\ &= \int_{30}^{33} \frac{1}{\ln(33/30)} dx \\ &= \frac{3}{\ln(33/30)}. \end{aligned}$$

### 5.3 Joint Probability Distributions for More Than Two Random Variables

**Definition.** A **joint probability density function** for the continuous random variables  $X_1, X_2, \dots, X_p$ , denoted as  $f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p)$ , satisfies the following properties:

- (1)  $f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p) \geq 0$
- (2)  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p) dx_1 dx_2 \dots dx_p = 1$
- (3) For any region  $B$  of  $p$ -dimensional space,

$$P((x_1, x_2, \dots, x_p) \in B) = \int \dots \int_B f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p) dx_1 \dots dx_p$$

**Definition.** The marginal density of  $X_1$  for joint probability density function

$$f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p)$$

is defined as

$$f_{X_1}(x_1) = \int \dots \int_{\bigcup_{k=2}^p X_k} f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p) dx_2 dx_3 \dots dx_p.$$

**Definition.** From above definition, we can define mean and variance that is

$$\begin{aligned} E(X_1) &= \int \dots \int_{\bigcup_{k=2}^p X_k} x_1 f_{X_1 X_2 \dots X_p}(x_1, x_2, \dots, x_p) dx_2 dx_3 \dots dx_p \\ &= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 \end{aligned}$$

and

$$Var(X_1) = \int_{-\infty}^{\infty} [x_1 - E(X_1)]^2 f_{X_1}(x_1) dx_1.$$

**Remark. Conditional probability distributions** can be developed for multiple random variables by an extension of the ideas used for two random variables. for example, the joint conditional probability distribution of  $X_1, X_2$  and  $X_3$  given  $X_4 = x_4$  and  $X_5 = x_5$  is

$$f_{X_1 X_2 X_3 | x_4 x_5}(x_1, x_2, x_3) = \frac{f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4 X_5}(x_4, x_5)}$$

for  $f_{X_4 X_5}(x_4, x_5) > 0$ .

**Example.** Suppose that the random variables  $X$ ,  $Y$  and  $Z$  have joint probability density function

$$f_{XYZ}(x, y, z) = c$$

over the cylinder  $x^2 + y^2 < 4$  and  $0 < z < 4$ .

- (1) Determine the constant  $c$  so that  $f_{XYZ}(x, y, z)$  is a probability distribution function.
- (2) Find  $P(X^2 + Y^2 < 2)$
- (3) Find  $P(Z < 2)$
- (4) Find  $E(X)$
- (5) Find  $P(X < 1|Y = 1)$
- (6) Find  $P(X^2 + Y^2 < 1|Y = 1)$
- (7) Find the conditional probability distribution of  $Z$  given that  $X = 1$  and  $Y = 1$ .

**Sol.**

- (1) We want to find  $c$  such that  $\iiint_T c \, dx \, dy \, dz = 1$ .

Since we know that it is over the cylinder  $x^2 + y^2 < 4$  and  $0 < z < 4$ .

Let  $T = \{(x, y, z) \mid x^2 + y^2 < 4, 0 < z < 4\}$ . This implies

$$\begin{aligned} \iiint_T c \, dx \, dy \, dz &= c \iiint_T 1 \, dx \, dy \, dz \\ &= c \iiint_T 1 \, dV \\ &= c \times 2^2 \times 4 \times \pi \\ &= 16\pi \times c = 1 \end{aligned}$$

Hence  $c = \frac{1}{16\pi}$ .

- (2) Let  $T = \{(x, y, z) \mid 0 < z < 4, -\sqrt{2} < y < \sqrt{2}, -\sqrt{2-y^2} < x < \sqrt{2-y^2}\}$ .

$$\begin{aligned} P(X^2 + Y^2 < 2) &= \iiint_T c \, dx \, dy \, dz \\ &= c \int_0^4 \int_{-\sqrt{2}}^{\sqrt{2}} \int_{-\sqrt{2-y^2}}^{\sqrt{2-y^2}} 1 \, dx \, dy \, dz \\ &= c \int_0^4 dz \int_{-\sqrt{2}}^{\sqrt{2}} 2\sqrt{2-y^2} \, dy \\ &= 8c \int_{-\sqrt{2}}^{\sqrt{2}} \sqrt{2-y^2} \, dy \\ &= 8 \times \frac{1}{16\pi} \times \pi = \frac{1}{2}. \end{aligned}$$

(3)

$$P(Z < 2) = c \iiint_T dV \times \frac{1}{2} = \frac{1}{2}.$$

(4) Let  $T = \{(x, y, z) \mid 0 < z < 4, -\sqrt{4-x^2} < y < \sqrt{4-x^2}, -2 < x < 2\}$ .

$$\begin{aligned} E(X) &= \int_{-2}^2 x f_X(x) dx \\ &= \int_{-2}^2 x \left[ \int_0^4 \int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} c dy dz \right] dx \\ &= 4c \int_{-2}^2 x \cdot 2\sqrt{4-x^2} dx \\ &= 8c \int_{-2}^2 x\sqrt{4-x^2} dx \\ &= 8c \left[ \frac{-1}{3} u^{3/2} \right]_0^4 = 0. \end{aligned}$$

(5) Let  $R = \{(x, z) \mid 0 < z < 4, -\sqrt{4-y^2} < x < \sqrt{4-y^2}\}$ .  
Compute

$$\begin{aligned} f_{XZ|Y=1}(x, z) &= \frac{c}{f_Y(y=1)} \\ &= c \left/ \iint_R c dx dz \right|_{y=1} \\ &= \left[ \int_0^4 \int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} 1 dx dz \right]_{y=1}^{-1} \\ &= \left[ \int_0^4 \int_{-\sqrt{3}}^{\sqrt{3}} 1 dx dz \right]^{-1} \\ &= \frac{1}{8\sqrt{3}} \end{aligned}$$

Since it is given  $Y = 1$ , we have

$$x^2 + 1 < 4 \iff -\sqrt{3} < x < \sqrt{3}.$$

Suppose  $\Omega = \{(x, z) \mid -\sqrt{3} < x < 1, 0 < z < 4\}$ .

And thus

$$\begin{aligned} P(X < 1 | Y = 1) &= \iint_{\Omega} f_{XZ|Y=1}(x, z) dx dz \\ &= \int_0^4 \int_{-\sqrt{3}}^1 \frac{1}{8\sqrt{3}} dx dz \\ &= 4 \times \frac{1 + \sqrt{3}}{8\sqrt{3}} = \frac{1 + \sqrt{3}}{2\sqrt{3}}. \end{aligned}$$

(6) Given  $x^2 + y^2 < 1$  and  $Y = 1$ , we obtain  $x^2 + 1 < 1 \iff x^2 < 0$ . Since there is no real number  $x$  satisfying this inequality, the probability is 0. Hence

$$P(X^2 + Y^2 < 1 | Y = 1) = 0.$$

(7) Compute

$$P_{Z|x=1, y=1}(z) = \frac{c}{f_{XY}(x=1, y=1)} = \frac{1}{4}.$$

## 5.4 Common Joint Distributions

### 5.4.1 Multinomial Probability Distribution

**Remark.** Recall binomial distribution, let  $X$  be the number of trials that result in a success  $x = 0, 1, \dots, n$ ,  $n = 1, 2, 3, \dots$  with the parameter  $p \in (0, 1)$ . Then

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

**Definition.** Suppose that a random experiment consists of a series of  $n$  trials. Assume that

- (1) The result of each trial is classified into one of  $k$  classes.
- (2) The probability of a trial generating a result in class 1, class 2,  $\dots$ , class  $k$  is  $p_1, p_2, \dots, p_k$ , respectively, and these probabilities are constant over the trials.
- (3) The trials are independent.

The random variables  $X_1, X_2, \dots, X_k$  that denote the number of trials that result in class 1, class 2,  $\dots$ , class  $k$ , respectively, have a **multinomial distribution**(多項分布) and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

for  $x_1 + x_2 + \dots + x_k = n$  and  $p_1 + p_2 + \dots + p_k = 1$ .

**Theorem.** If  $X_1, X_2, \dots, X_k$  have a multinomial distribution, the marginal probability distribution of  $X_i$  is binomial with

$$E(X_i) = np_i$$

and

$$V(X_i) = np_i(1 - p_i).$$

### 5.4.2 Bivariate Normal Probability Density Function

**Remark.** Recall normal distribution, we know

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where  $-\infty < x < \infty$ .

**Definition.** The probability density function of a **bivariate normal distribution**(二變數常態分佈) is

$$f_{XY}(x, y, ) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^u$$

for

$$u = \frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right],$$

where

$$\begin{aligned} -\infty < x < \infty, & & -\infty < y < \infty, \\ \sigma_X > 0, & & -\infty < \mu_X < \infty, \\ \sigma_Y > 0, & & -\infty < \mu_Y < \infty, \\ -1 < \rho < 1. & & \end{aligned}$$

**Theorem.** Suppose  $X$  and  $Y$  are bivariate normal distribution. If  $\rho = 0$ , then  $X$  and  $Y$  are independent.

In other view, we can use a matrix to define bivariate normal distribution and even more generalizing.

**Definition.** The probability density function of a multivariate normal distribution is

$$f_X(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} e^{\left( -\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \right)}$$

for  $\Sigma$  is a covariance matrix of  $X_1, \dots, X_k$ .

**Remark.** In bivariate case, it is defined

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

## 5.5 Linear Function of Random Variables

**Definition.** Suppose  $X_1, X_2, \dots, X_p$  are random variables and  $c_1, c_2, \dots, c_p \in \mathbb{R}$ . Then the linear combination

$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$$

is a linear function of  $X_1, X_2, \dots, X_p$ .

**Remark.** If  $Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$ , then

$$(1) E(Y) = c_1E(X_1) + c_2E(X_2) + \dots + c_pE(X_p)$$

(2)

$$\begin{aligned} Var(Y) &= c_1^2Var(X_1) + c_2^2Var(X_2) + \dots + c_p^2Var(X_p) \\ &\quad + 2 \sum_{i < j} c_i c_j cov(X_i, X_j) \end{aligned}$$

**Theorem.** If  $X_1, X_2, \dots, X_p$  are independent such that  $cov(X_i, X_j) = 0$  for  $i \neq j$ , then

$$Var(Y) = c_1^2Var(X_1) + c_2^2Var(X_2) + \dots + c_p^2Var(X_p).$$

**Remark.** Let  $\bar{X} = \frac{X_1 + X_2 + \dots + X_p}{p}$  be the average of random variable.

(1)  $E(\bar{X}) = \frac{1}{p}E(X_1 + X_2 + \dots + X_p)$ , if  $X_i$  are independent and  $E(X_i) = \mu$ , we obtain  $E(\bar{X}) = \mu$ .

(2) Suppose  $X_i$  are independent and  $Var(X_i) = \sigma^2$ , then

$$Var(\bar{X}) = \frac{\sigma^2}{p}.$$

*Proof.* (1) Suppose  $X_i$  are independent and  $E(X_i) = \mu$ , thus

$$\begin{aligned} E(\bar{X}) &= \frac{1}{p}E(X_1 + X_2 + \dots + X_p) \\ &= \frac{1}{p} \sum_{k=1}^p E(X_k) \\ &= \frac{1}{p} \sum_{k=1}^p \mu = \frac{1}{p} \times p\mu = \mu. \end{aligned}$$

(2)

$$Var(\bar{X}) = \frac{1}{p^2} \sum_{k=0}^p Var(X_k) = \frac{1}{p^2} \cdot p\sigma^2 = \frac{\sigma^2}{p}.$$

□

## 5.6 General Functions of Random Variables

Let  $y = f(x)$ , we can define a transformation from  $X$  to  $Y$ .

**Definition.** Suppose that  $X$  is a **discrete** random variable with probability distribution  $f_X(x)$ . Let  $Y = h(X)$  define a one-to-one transformation between the values of  $X$  and  $Y$  so that the equation  $y = h(x)$  can be solved uniquely for  $x$  in terms of  $y$ . Then the probability mass function of the random variable  $Y$  is

$$f_Y(y) = f_X[u(y)].$$

**Example.** Let  $X$  be a geometric random variable with pmf

$$f_X(x) = p(1-p)^{x-1}, \quad x = 1, 2, \dots$$

Find the pmf of  $Y = X^2$ .

**Sol.** Let  $f_Y(y) = P(Y = y)$ , since  $Y = X^2$ , it implies  $X = \sqrt{Y}$ . Thus

$$P(Y = y) = P(X = \sqrt{y}).$$

Because  $P(X = x) = f_X(x) = p(1-p)^{x-1}$ ,  $x = 1, 2, \dots$ , we have

$$P(Y = y) = P(X = \sqrt{y}) = p(1-p)^{\sqrt{y}-1} = f_Y(y).$$

Hence

$$f_Y(y) = p(1-p)^{\sqrt{y}-1}, \quad y = 1^2, 2^2, 3^2, \dots$$

Suppose that  $X$  is a **continuous** random variable with probability distribution  $f_X(x)$ . Let  $Y = g(X)$  be a random variable and  $g(x)$  is strictly monotonic and differentiable. Since  $F'_Y(y) = f_Y(y)$ , it is clear that  $F_Y(y) = P(Y \leq y)$ . Therefore, because  $Y = g(X)$  and  $g^{-1}(Y)$  exists, assume  $g^{-1}$  is decreasing, we obtain  $P(Y \leq y) = P(g(x) \leq y)$ . Then

$$\begin{aligned} F_y(y) &= P(g(x) \leq y) \\ &= P(g^{-1}(g(x)) \leq g^{-1}(y)) \\ &= P(X \geq g^{-1}(y)) \\ &= 1 - P(X \leq g^{-1}(y)) \\ &= 1 - F_X(g^{-1}(y)) \end{aligned}$$

Differentiate  $F_y(y)$ , we have

$$\begin{aligned} F'_Y(y) &= 0 - f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} \\ &= -f_X(g^{-1}(y)) \cdot (-1) \left| \frac{dg^{-1}(y)}{dy} \right| \\ &= f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|. \end{aligned}$$

**Theorem.** Given  $X$  is a **continuous** random variable with probability distribution  $f_X(x)$ . Let  $Y = g(X)$  be a random variable and  $g(x)$  is strictly monotonic and differentiable. Then

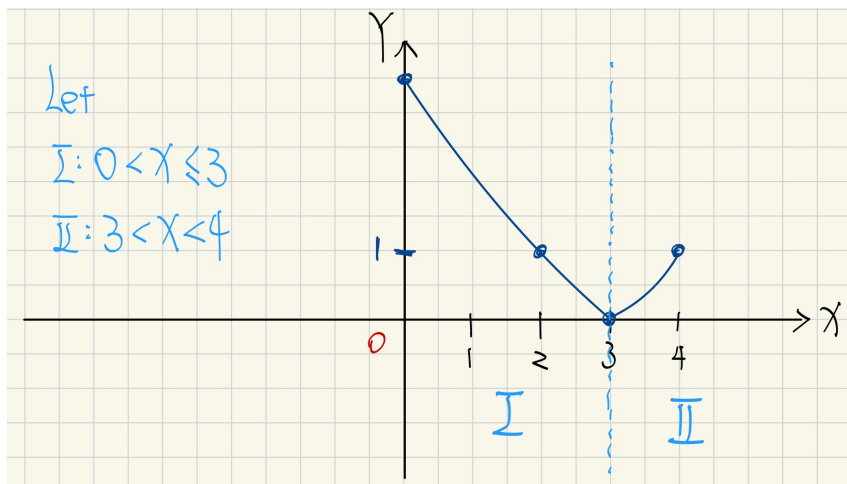
$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|.$$

**Remark.** In continuous,  $f_X(x) \neq P(X = x)$ . (不一定, 根據連續的 cdf 和 pdf 的特性)

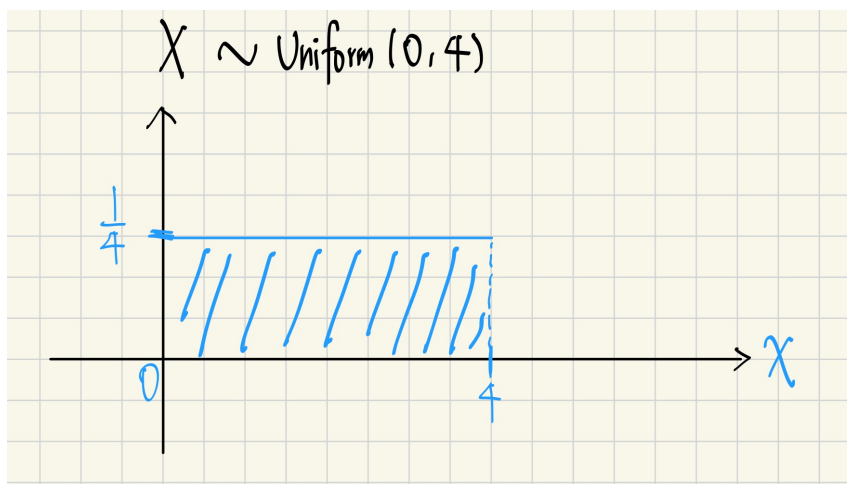
**Example.** Let  $X$  be uniformly distributed RV over  $(0, 4)$ . Let  $Y = g(X) = (X-3)^2$ . Find the pdf of  $Y$ .

**Sol.** Consider:

(1)  $Y = (X - 3)^2 = g(X)$



(2) Since it is uniform, we have



Hence

$$f_X(x) = \begin{cases} \frac{1}{4}, & 0 < x < 4 \\ 0, & \text{otherwise} \end{cases}$$

(I) For  $0 < X \leq 3$ ,  $Y = (X - 3)^2 \implies Y = g(X)$

$$\begin{aligned}\pm\sqrt{Y} = X - 3 &\iff X = 3 \pm \sqrt{Y} \\ &\iff X = 3 - \sqrt{Y} = g^{-1}(Y).\end{aligned}$$

(II) For  $3 < X < 4$ , it is

$$X = 3 + \sqrt{Y} = g^{-1}(Y).$$

Using

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|,$$

we have

(I) For  $0 < X \leq 3$ ,

$$\begin{aligned}f_{Y_I}(y) &= f_X(3 - \sqrt{Y}) \left| \frac{d}{dy}(3 - \sqrt{Y}) \right| \\ &= \frac{1}{4} \cdot \left| -\frac{1}{2\sqrt{Y}} \right|, \quad 0 < y < 9.\end{aligned}$$

(II) For  $3 < X < 4$ ,

$$\begin{aligned}f_{Y_{II}}(y) &= \frac{1}{4} \cdot \left| \frac{d}{dy}(3 + \sqrt{Y}) \right| \\ &= \frac{1}{4} \cdot \left| \frac{1}{2\sqrt{Y}} \right|, \quad 0 < y < 1.\end{aligned}$$

Hence

$$\begin{aligned}f_Y(y) &= \begin{cases} f_{Y_I}(y) + f_{Y_{II}}(y), & 0 < y < 1 \\ f_{Y_I}(y), & 1 < y < 9 \end{cases} \\ &= \begin{cases} \frac{1}{4\sqrt{y}}, & 0 < y < 1 \\ \frac{1}{8\sqrt{y}}, & 1 < y < 9. \end{cases}\end{aligned}$$

## 5.7 General Functions of Bivariate Random Variables

Let  $X$  and  $Y$  be random variables, and defined transformation

$$\begin{cases} U = g(X, Y) \\ V = h(X, Y). \end{cases}$$

Then the inverse transformation is

$$\begin{cases} X = h_1(U, V) \\ Y = h_2(U, V). \end{cases}$$

Given  $f_{XY}(x, y)$ , then we define

$$f_{UV}(u, v) = f_{XY}(h_1, h_2)|J|$$

where  $J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \neq 0$  is a Jacobian.

**Example.** Consider the following transformation

$$T : \begin{cases} u = 3x + 2y \\ v = x - y \end{cases}$$

- (a) Is the transformation invertible? If yes, find  $T^{-1}$ .
- (b) Find the Jacobian for  $T^{-1}$ .

**Sol.** (a) Compute  $T^{-1}$ , we have

$$T^{-1} : \begin{cases} x = \frac{u + 2v}{5} \\ y = \frac{u - 3v}{5} \end{cases}$$

(b) Compute

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & -\frac{3}{5} \end{vmatrix} \\ &= \frac{1}{25}(-3 - 2) = -\frac{1}{5} \end{aligned}$$

**Example.** Let  $X$  and  $Y$  be independent uniformly distributed random variables over the intervals  $(0, 2)$  and  $(0, 3)$ , respectively. Let  $U = XY$  and  $V = Y$ , i.e.

$$T : \begin{cases} u = xy \\ v = y \end{cases}$$

(a) Find  $f_X(x)$ ,  $f_Y(y)$  and  $f_{XY}(x, y)$ .

(b) Find Jacobian.

(c) Find  $f_{UV}(u, v)$

**Sol.** (a) Since  $X \sim \text{uniform}(0, 2)$  and  $Y \sim \text{uniform}(0, 3)$ , therefore

$$\begin{cases} f_X(x) = \frac{1}{2}, & 0 < x < 2 \\ f_Y(y) = \frac{1}{3}, & 0 < y < 3. \end{cases}$$

Since  $X$  and  $Y$  are independent, then

$$f_{XY}(x, y) = \frac{1}{6}, \quad 0 < x < 2, \quad 0 < y < 3.$$

(b)

$$T : \begin{cases} u = xy \\ v = y \end{cases} \implies T^{-1} : \begin{cases} x = \frac{u}{v} \\ y = v \end{cases}$$

Compute

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{v} & \frac{-u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}, \quad 0 < v < 3.$$

(c) By the formula  $f_{UV}(u, v) = f_{XY}(h_1, h_2)|J|$ , hence

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{6} \cdot \left| \frac{1}{v} \right| \\ &= \frac{1}{6v}, \quad 0 < u < 2v, \quad 0 < v < 3 \end{aligned}$$

By

$$\begin{aligned} 0 < x < 2 &\iff 0 < \frac{u}{v} < 2 \\ &\iff 0 < u < 2v \\ &\iff \left( 0 < \frac{u}{2} < v < 3 \right). \end{aligned}$$