Core course for *App.AI*, *Bioinfo.*, *Data Sci.&Eng.*, *Dec.Analytics*, *Q.Fin.*, *Risk Mgmt.* and *Stat.* Majors:

STAT2601A Probability and Statistics I (2023-2024 First Semester)

Chapter 7: Multivariate Distributions

7.1 Joint and Marginal Distributions

When an experiment or survey is conducted, two or more random variables are often observed simultaneously not only to study their individual probabilistic behaviours but also to determine the degree of relationship among the variables as in most of cases, the variables are related. The probabilistic behaviours of the random variables are described by their *joint distribution*.

In the simplest case, suppose there are only two discrete random variables X and Y (sometimes denoted by a random vector (X, Y)) which take distinct values:

$$\begin{aligned} &\text{Support of } X: \quad X(\Omega) = \{x_1, x_2, \ldots, x_r\} \\ &\text{Support of } Y: \quad Y(\Omega) = \{y_1, y_2, \ldots, y_c\} \end{aligned}$$

The joint distribution of two random variables is called a *bivariate distribution*.

Definition 7.1.
The *joint probability mass function* (*joint pmf*) of the discrete random variables X and Y and is defined by

$$p(x, y) = \Pr(X = x, Y = y), \qquad x \in X(\Omega),\ y \in Y(\Omega).$$

◀

Sometimes the joint pmf can be conveniently presented in the form of a two-way table as

Value of X	Value of Y			
	y_1	y_2	$\cdots$	y_c
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_c)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_c)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_r	$p(x_r, y_1)$	$p(x_r, y_2)$	$\cdots$	$p(x_r, y_c)$

Example 7.1.

Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls. If we let X and Y denote, respectively, the number of red and white balls in the sample, then both X and Y takes values $0, 1, 2, 3$ only. The joint pmf of (X, Y) can be calculated as

$$
\begin{aligned}
p(0,0) &= \Pr(X=0, Y=0) = \Pr(3 \text{ blue balls}) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}, \\
p(0,1) &= \Pr(X=0, Y=1) = \Pr(1 \text{ white ball, } 2 \text{ blue balls}) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}, \\
p(2,1) &= \Pr(X=2, Y=1) = \Pr(2 \text{ red balls, } 1 \text{ white ball}) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220}, \\
p(2,2) &= \Pr(X=2, Y=2) = \Pr(2 \text{ red balls, } 2 \text{ white balls}) = 0.
\end{aligned}
$$

Based on similar calculations, we have

Value of X	Value of Y				Total
	0	1	2	3	
0	0.0454	0.1818	0.1364	0.0182	0.3818
1	0.1364	0.2727	0.0818	0	0.4909
2	0.0682	0.0545	0	0	0.1227
3	0.0045	0	0	0	0.0045
Total	0.2545	0.5091	0.2182	0.0182	1.0000

The above probabilities can be also represented by the following expression:

$$
p(x,y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}}, \qquad x = 0,1,2,3, \; y = 0,1,2,3, \; x+y \le 3.
$$

It is called the *bivariate hypergeometric distribution.*

★

Conditions for a joint pmf

1. $0 \le p(x,y) \le 1$ for all $x \in X(\Omega)$ and $y \in Y(\Omega)$.
2. $\displaystyle\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p(x,y) = 1.$
3. $\displaystyle\Pr((X,Y) \in A) = \sum_{(x,y) \in A} p(x,y)$ where $A \subset X(\Omega) \times Y(\Omega)$.

Example 7.2.

For the joint pmf in **Example 7.1.**, obviously p satisfies properties 1 and 2.

For the probability that there are same number of red and white balls,

$$\begin{aligned}\Pr(X = Y) &= \Pr((X, Y) \in \{(0, 0), (1, 1), (2, 2), (3, 3)\}) \\ &= p(0, 0) + p(1, 1) \\ &= 0.0454 + 0.2727 = 0.3181.\end{aligned}$$

For the probability that there are less red balls than white balls,

$$\begin{aligned}\Pr(X < Y) &= \Pr((X, Y) \in \{(0, 1), (0, 2), (0, 3), (1, 2), (1, 3), (2, 3)\}) \\ &= p(0, 1) + p(0, 2) + p(0, 3) + p(1, 2) \\ &= 0.1818 + 0.1364 + 0.0182 + 0.0818 = 0.4182.\end{aligned}$$

We may also compute the probability concerning X only. For example,

$$\begin{aligned}\Pr(X = 0) &= p(0, 0) + p(0, 1) + p(0, 2) + p(0, 3) \\ &= 0.0454 + 0.1818 + 0.1364 + 0.0182 = 0.3818. \\ \Pr(X = 1) &= p(1, 0) + p(1, 1) + p(1, 2) + p(1, 3) \\ &= 0.1364 + 0.2727 + 0.0818 + 0 = 0.4909.\end{aligned}$$

★

As can be seen, the probabilistic behaviour of X (or Y) alone can be obtained directly from the joint distribution of X and Y. The probability of each value is the corresponding row (column) sum. Since the probabilities obtained are the marginal totals from the two-way table, the distribution of X (Y) alone is called the *marginal distribution* of X (Y). Thus we have the following definition.

Definition 7.2.
Let X and Y be discrete random variables with joint pmf $p(x, y)$. The marginal pmfs of X and Y are respectively defined as

$$p_X(x) = \Pr(X = x) = \sum_{y \in Y(\Omega)} p(x, y),$$

and

$$p_Y(y) = \Pr(Y = y) = \sum_{x \in X(\Omega)} p(x, y).$$

◀

Example 7.3.

For the (X, Y) in **Example 7.1.** and **Example 7.2.**, the marginal pmf of X is given by

$$p_X(x) = \Pr(X = x) = \begin{cases} 0.3818, & x = 0; \\ 0.4909, & x = 1; \\ 0.1227, & x = 2; \\ 0.0045, & x = 3. \end{cases}$$

The marginal pmf of Y is given by

$$p_Y(y) = \Pr(Y = y) = \begin{cases} 0.2545, & y = 0; \\ 0.5091, & y = 1; \\ 0.2182, & y = 2; \\ 0.0182, & y = 3. \end{cases}$$

★

Remark

Joint pmf can uniquely determine the marginal pmfs, but the converse is not true.

Example 7.4.

The following table shows a different joint pmf from **Example 7.3.** that yields the same marginal pmfs.

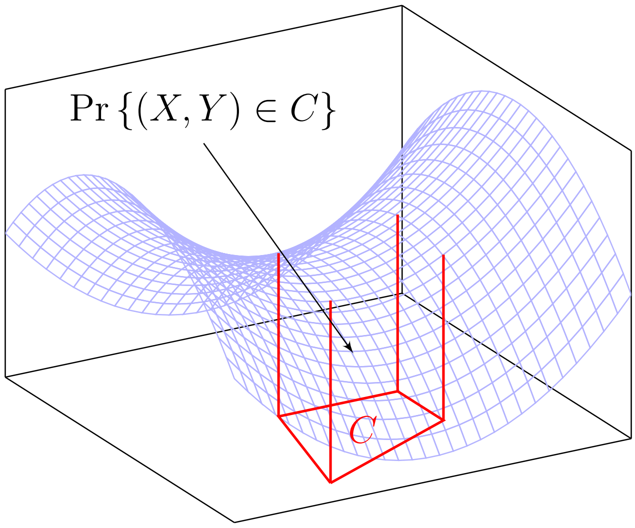
Value of X	Value of Y				Total
	0	1	2	3	
0	0.0972	0.1944	0.0833	0.0069	0.3818
1	0.1249	0.2499	0.1071	0.0089	0.4909
2	0.0312	0.0625	0.0268	0.0022	0.1227
3	0.0011	0.0023	0.0010	0.0001	0.0045
Total	0.2545	0.5091	0.2182	0.0182	1.0000

★

Definition 7.3.
We say that X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every (*measurable*) set C in the two-dimensional plane,

$$\Pr\{(X, Y) \in C\} = \iint_{(x,y)\in C} f(x, y)\mathrm{d}x\mathrm{d}y.$$

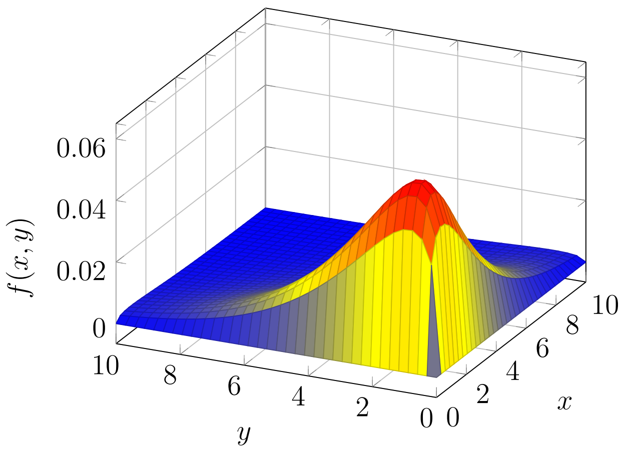
This function, if exists, is called the *joint probability density function* (*joint pdf*). ◀

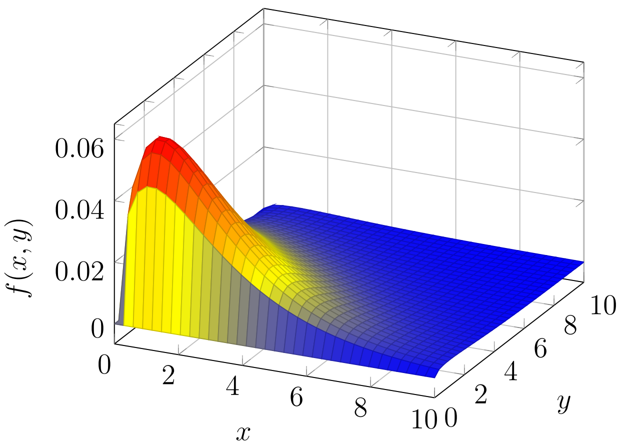


Example 7.5. (Bivariate chi-squared distribution)

The joint probability density function is

$$f(x, y) = \frac{1}{2\pi}(xy)^{\frac{1}{2}} \exp\left(-\frac{x+y}{2}\right), \qquad 0 < x, y < \infty.$$



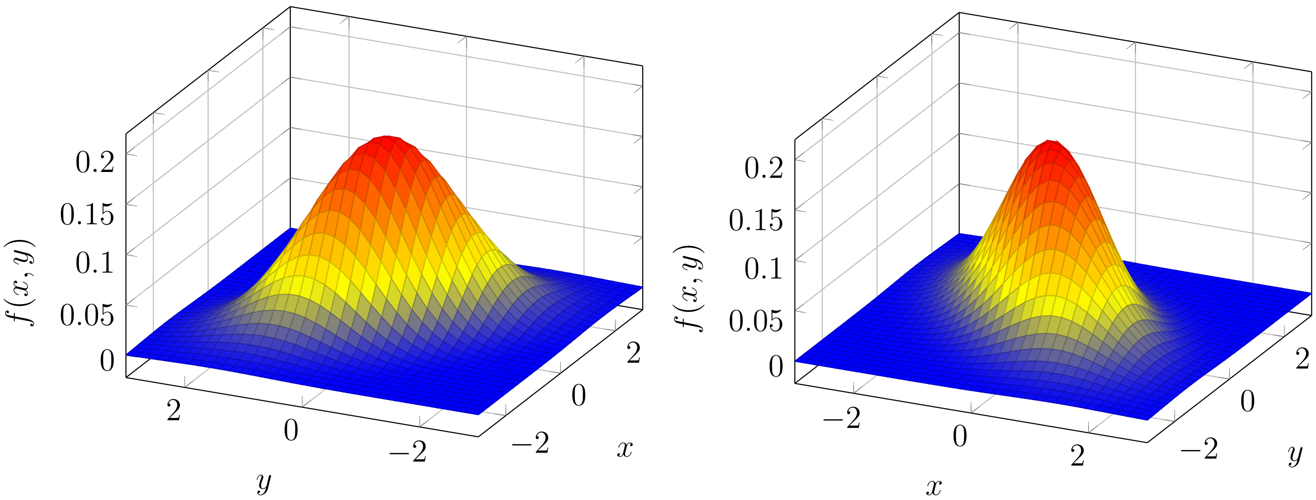


★

Example 7.6. (Bivariate normal distribution)

The joint probability density function is

$$f(x,y) = \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{2}{3}(x^2 + xy + y^2)\right\}, \qquad x, y \in \mathbb{R}.$$



Properties of joint pdf

1. $f(x,y) \geq 0$ for all $-\infty < x < \infty$ and $-\infty < y < \infty$.
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = 1$.
3. $\Pr\left((X,Y) \in A\right) = \iint_A f(x,y)\mathrm{d}x\mathrm{d}y$.
   In particular, $\Pr(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\mathrm{d}x\mathrm{d}y$.
4. *Joint distribution function*:
$$F(x,y) = \Pr(X \leq x, Y \leq y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)\mathrm{d}s\mathrm{d}t.$$
5. *Marginal pdf*:
$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y, \qquad -\infty < x < \infty, \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x, \qquad -\infty < y < \infty. \end{aligned}$$

Example 7.7.

Consider $f(x, y) = 4x(1 - y)$, $0 \le x \le 1$ and $0 \le y \le 1$.

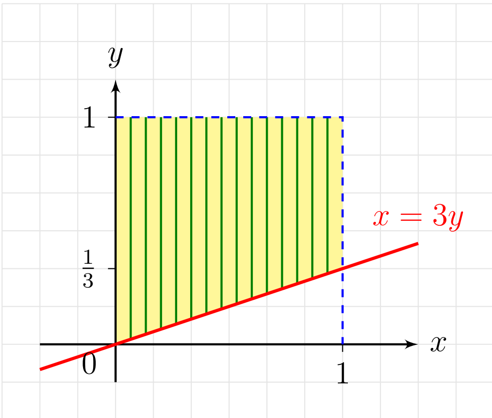
It is easy to verify that f is a joint pdf as $f(x, y) \ge 0$ for all $0 \le x \le 1$, $0 \le y \le 1$ and

$$\int_0^1 \int_0^1 4x(1-y)\mathrm{d}y\mathrm{d}x = \int_0^1 4x\left[y - \frac{y^2}{2}\right]_0^1 \mathrm{d}x = \int_0^1 2x\mathrm{d}x = \left[x^2\right]_0^1 = 1.$$

Similarly,

$$\Pr\left(0 \le X \le \frac{1}{2}, \frac{1}{2} \le Y \le 1\right) = \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 4x(1-y)\mathrm{d}y\mathrm{d}x = \frac{1}{16}.$$

Suppose we want to determine $\Pr(X < 3Y)$. The following graph shows the region corresponding to the event $X < 3Y$.



The points in the region can be specified by $y > \frac{x}{3}$, $0 < x < 1$, therefore we have

$$\begin{aligned}
\Pr(X < 3Y) &= \int_0^1 \int_{\frac{x}{3}}^1 4x(1-y)\mathrm{d}y\mathrm{d}x \\
&= \int_0^1 4x\left[y - \frac{y^2}{2}\right]_{\frac{x}{3}}^1 \mathrm{d}x \\
&= \int_0^1 4x\left(\frac{1}{2} - \frac{x}{3} + \frac{x^2}{18}\right)\mathrm{d}x \\
&= \int_0^1 \left(2x - \frac{4x^2}{3} + \frac{2x^3}{9}\right)\mathrm{d}x \\
&= \left[x^2 - \frac{4x^3}{9} + \frac{x^4}{18}\right]_0^1 \\
&= 1 - \frac{4}{9} + \frac{1}{18} = \frac{11}{18}.
\end{aligned}$$

For the joint distribution function, we need to consider different ranges.

For $0 \le x \le 1$, $0 \le y \le 1$,

$$F(x,y) = \int_0^x \int_0^y 4s(1-t)\mathrm{d}t\mathrm{d}s = x^2(2y-y^2).$$

For $0 \le x \le 1$, $y > 1$,

$$F(x,y) = F(x,1) = x^2.$$

For $x > 1$, $0 \le y \le 1$,

$$F(x,y) = F(1,y) = 2y - y^2.$$

Therefore,

$$F(x,y) = \begin{cases} 0, & x < 0 \text{ or } y < 0; \\ x^2(2y-y^2), & 0 \le x \le 1,\ 0 \le y \le 1; \\ x^2, & 0 \le x \le 1,\ y > 1; \\ 2y - y^2, & x > 1,\ 0 \le y \le 1; \\ 1, & x > 1,\ y > 1. \end{cases}$$

The marginal pdfs are

$$\begin{aligned} f_X(x) &= \int_0^1 4x(1-y)\mathrm{d}y = 4x\left[y - \frac{y^2}{2}\right]_0^1 = 2x, \qquad 0 \le x \le 1, \\ f_Y(y) &= \int_0^1 4x(1-y)\mathrm{d}x = (1-y)\left[2x^2\right]_0^1 = 2(1-y), \qquad 0 \le y \le 1. \end{aligned}$$

★

7.2 Independence of Random Variables

The concept of the independence among events described in **Chapter 3** can be generalized to the independence among random variables.

Definition 7.4.
Two random variables X and Y are said to be *independent* if and only if their joint pmf (pdf) is equal to the product of their marginal pmfs (pdfs), i.e.,

$$p(x,y) = p_X(x)p_Y(y), \qquad \text{for all } x, y, \qquad \text{if } X, Y \text{ are discrete;}$$

or

$$f(x,y) = f_X(x)f_Y(y), \qquad \text{for all } x, y, \qquad \text{if } X, Y \text{ are continuous.}$$

◀

Thus X and Y are dependent if there exists x and y such that

$$p(x,y) \ne p_X(x)p_Y(y);$$

or

$$f(x,y) \ne f_X(x)f_Y(y).$$

Example 7.8.

In **Example 7.1.**, X is the number of red balls, Y is the number of white balls in a sample of 3 randomly drawn from an urn containing 3 red balls, 4 white balls, and 5 blue balls. The following table shows the joint and marginal pmfs.

Value of X	Value of Y				Total
	0	1	2	3	
0	0.0454	0.1818	0.1364	0.0182	0.3818
1	0.1364	0.2727	0.0818	0	0.4909
2	0.0682	0.0545	0	0	0.1227
3	0.0045	0	0	0	0.0045
Total	0.2545	0.5091	0.2182	0.0182	1.0000

Note that

$$p_X(0)p_Y(0) = 0.3818 \times 0.2545 = 0.0972 \neq 0.0454 = p(0,0).$$

Therefore, X and Y are dependent, i.e., knowing the value of X will affect the uncertainty about Y, and vice versa.

★

Example 7.9.

In **Example 7.7.**, the joint pdf is $f(x,y) = 4x(1-y)$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$, and the marginal pdfs are given by

$$f_X(x) = 2x, \quad 0 \leq x \leq 1, \qquad \text{and} \qquad f_Y(y) = 2(1-y), \quad 0 \leq y \leq 1.$$

Since $f_X(x)f_Y(y) = 4x(1-y) = f(x,y)$ for all $0 \leq x \leq 1$ and $0 \leq y \leq 1$, X and Y are independent.

★

Example 7.10.

Consider the following joint pmf $p_{X,Y}(x, y)$.

x	y				$p_X(x)$
	10	20	40	80	
20	$0.04 = (0.2)(0.2)$	$0.06 = (0.2)(0.3)$	$0.06 = (0.2)(0.3)$	$0.04 = (0.2)(0.2)$	0.2
40	$0.10 = (0.5)(0.2)$	$0.15 = (0.5)(0.3)$	$0.15 = (0.5)(0.3)$	$0.10 = (0.5)(0.2)$	0.5
60	$0.06 = (0.3)(0.2)$	$0.09 = (0.3)(0.3)$	$0.09 = (0.3)(0.3)$	$0.06 = (0.3)(0.2)$	0.3
$p_Y(y)$	0.2	0.3	0.3	0.2	1

Since the products $p_X(x)p_Y(y)$ agrees everywhere with $p(x, y)$, the random variables X and Y are independent.

On the other hand, consider the following pmf.

x	y				$p_X(x)$
	10	20	40	80	
20	0.04	0.06	0.06	0.04	0.2
40	0.10	0.15	0.15	0.05	0.45
60	0.06	0.09	0.09	0.11	0.35
$p_Y(y)$	0.2	0.3	0.3	0.2	1

Although we have $p(20, 10) = p_X(20)p_Y(10)$, $p(20, 20) = p_X(20)p_Y(20)$, ..., the random variables X and Y are not independent as $p(40, 80) \neq p_X(40)p_Y(80)$.

★

Theorem 7.1.

Let X and Y be random variables with joint pdf (or pmf) $f(x, y)$. Then X and Y are independent if and only if

(i) the supports of X and Y do not depend on each other (i.e., the region of possible values is a rectangle); and

(ii) $f(x, y)$ can be factorized as $g(x)h(y)$.

This proposition also applies to discrete random variables.

▲

Example 7.11.

Consider a joint pdf

$$f(x,y) = \begin{cases} \frac{1}{4}(x+1)(y+1)e^{-x-y}, & x,y>0; \\ 0, & \text{otherwise.} \end{cases}$$

X and Y are independent since the supports do not depend on each other and

$$f(x,y) = \left[\frac{1}{4}(x+1)e^{-x}\right]\left[(y+1)e^{-y}\right].$$

Note: Actually the marginal pdfs are

$$\begin{aligned} f_X(x) &= \frac{1}{2}(x+1)e^{-x} \qquad \text{for } x>0, \\ f_Y(y) &= \frac{1}{2}(y+1)e^{-y} \qquad \text{for } y>0. \end{aligned}$$

★

Example 7.12.

Consider a joint pdf

$$f(x,y) = \begin{cases} \frac{1}{2}(x+y)e^{-x-y}, & x,y>0; \\ 0, & \text{otherwise.} \end{cases}$$

X and Y are NOT independent as $f(x,y)$ cannot be factorized as $g(x)h(y)$.

★

Example 7.13.

Suppose we randomly choose a point uniformly within the unit circle $x^2+y^2=1$. Then the joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \le 1; \\ 0, & x^2+y^2 > 1. \end{cases}$$

Although $f(x,y)$ is a constant, X and Y are NOT independent because the support of X is from $-\sqrt{1-y^2}$ to $\sqrt{1-y^2}$ which depends on the value of y.

★

Remarks

1. The definitions of joint pdf (pmf) and marginal pdf (pmf) can be generalized to multivariate case directly.

   - Joint pmf of $X_1, X_2, \ldots, X_n$: $p(x_1, x_2, \ldots, x_n) = \Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$
   - Joint pdf of $X_1, X_2, \ldots, X_n$: $f(x_1, x_2, \ldots, x_n)$

- Marginal pmf/pdf of X_1:
  - Discrete
  $$p_{X_1}(x_1) = \sum_{x_2}\sum_{x_3}\cdots\sum_{x_n} p(x_1, x_2, \ldots, x_n), \qquad x_1 = X_1(\Omega).$$
  - Continuous
  $$f_{X_1}(x_1) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\mathrm{d}x_2\mathrm{d}x_3\cdots\mathrm{d}x_n, \qquad -\infty < x_1 < \infty.$$
- $X_1, X_2, \ldots, X_n$ are said to be (mutually) independent if and only if
  - Discrete
  $$p(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n), \qquad \text{for all } x_1, x_2, \ldots, x_n.$$
  - Continuous
  $$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n), \qquad \text{for all } x_1, x_2, \ldots, x_n.$$

2. If $X_1, X_2, \ldots, X_n$ are independent, then for all subsets $A_1, A_2, \ldots, A_n \subset (-\infty, \infty)$,
$$\Pr(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \Pr(X_1 \in A_1)\Pr(X_2 \in A_2)\cdots\Pr(X_n \in A_n).$$
In particular,
$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n), \qquad \text{for all } x_1, x_2, \ldots, x_n.$$
The converse is also true.

7.3 Expectation of Function of Random Variables

Definition 7.5.
For random variables $X_1, X_2, \ldots, X_n$ (not necessarily independent) with joint pmf $p(x_1, x_2, \ldots, x_n)$ or joint pdf $f(x_1, x_2, \ldots, x_n)$; if $u(X_1, X_2, \ldots, X_n)$ is a function of these random variables, then the expectation of $u(X_1, X_2, \ldots, X_n)$ is defined as

- Discrete
$$\mathrm{E}[u(X_1, X_2, \ldots, X_n)] = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} u(x_1, x_2, \ldots, x_n)p(x_1, x_2, \ldots, x_n).$$
- Continuous
$$\mathrm{E}[u(X_1, X_2, \ldots, X_n)] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} u(x_1, x_2, \ldots, x_n)f(x_1, x_2, \ldots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n.$$

◀

Example 7.14.

Pairs of resistors are to be connected in parallel and a difference in electrical potential applied across the resistor assembly. Ohm's law predicts that in such a situation, the combined resistance would be

$$R = \left(\frac{1}{R_1} + \frac{1}{R_2}\right)^{-1},$$

where R_1 and R_2 are the two resistances. Suppose that experience reveals that in a specific component of a machine the two resistances have a joint pdf

$$f(x,y) = \begin{cases} \frac{1}{2}(x+y)e^{-x-y}, & x>0,\ y>0; \\ 0, & \text{otherwise.} \end{cases}$$

Then the expected value of the combined resistance is given by

$$\begin{aligned} \mathrm{E}(R) &= \int_0^\infty \int_0^\infty \left(\frac{1}{x} + \frac{1}{y}\right)^{-1} f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2}\int_0^\infty \int_0^\infty xye^{-x-y}\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{2}\left(\int_0^\infty xe^{-x}\mathrm{d}x\right)\left(\int_0^\infty ye^{-y}\mathrm{d}y\right) \\ &= \frac{1}{2}(\Gamma(2))^2 \\ &= 0.5 \text{ (in unit of ohm } (\Omega)). \end{aligned}$$

If the two resistors are to be connected in series, then the combined resistance would be

$$R = R_1 + R_2.$$

Thought Question:

Can we compute $\mathrm{E}(R)$ by $\mathrm{E}(R_1) + \mathrm{E}(R_2)$?

★

Properties

1. $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$. ($X$ and Y need not be independent.)
   In general, $\mathrm{E}[u_1(X) + u_2(Y)] = \mathrm{E}[u_1(X)] + \mathrm{E}[u_2(Y)]$.
2. If X and Y are independent, then $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.
   In general, $\mathrm{E}[u_1(X)u_2(Y)] = \mathrm{E}[u_1(X)]\mathrm{E}[u_2(Y)]$.
   The converse is NOT necessarily true.
3. If X and Y are independent, then the moment generating function of $X+Y$ is equal to the product of the moment generating functions of X and Y, i.e.,
   $$M_{X+Y}(t) = M_X(t)M_Y(t),$$
   and also for the linear combination $aX+bY$, the moment generating function is given by
   $$M_{aX+bY}(t) = M_X(at)M_Y(bt).$$

4. If $u(X_1, X_2, \ldots, X_n) = g(X_1)$ (i.e., u is a function of X_1 only), then the expectation of g can be obtained by the marginal pmf/pdf of X_1. That is, (for discrete version)

$$\begin{aligned} \mathrm{E}[g(X_1)] &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} g(x_1)p(x_1, x_2, \ldots, x_n) \\ &= \sum_{x_1} g(x_1)\sum_{x_2}\sum_{x_3}\cdots\sum_{x_n} p(x_1, x_2, \ldots, x_n) \\ &= \sum_{x_1} g(x_1)p_{X_1}(x_1). \end{aligned}$$

The continuous version is similar.

Example 7.15.

For **Example 7.14.**, the marginal pdfs of R_1 and R_2 are

$$f_{R_1}(x) = \begin{cases} \frac{1}{2}(x+1)e^{-x}, & x > 0; \\ 0, & \text{otherwise.} \end{cases} \quad \text{and} \quad f_{R_2}(y) = \begin{cases} \frac{1}{2}(y+1)e^{-y}, & y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned} \mathrm{E}(R_1) &= \int_0^\infty \frac{1}{2}x(x+1)e^{-x}\mathrm{d}x = \frac{3}{2}, \\ \mathrm{E}(R_2) &= \mathrm{E}(R_1) = \frac{3}{2}, \\ \mathrm{E}(R) &= \mathrm{E}(R_1) + \mathrm{E}(R_2) = \frac{3}{2} + \frac{3}{2} = 3. \end{aligned}$$

Since $f_{R_1}(x)f_{R_2}(y) = \dfrac{1}{4}(x+1)(y+1)e^{-x-y} \neq f(x, y)$ for all $x > 0$ and $y > 0$, R_1 and R_2 are not independent. To compute $\mathrm{E}(R_1R_2)$, we may need to evaluate

$$\mathrm{E}(R_1R_2) = \int_0^\infty \int_0^\infty \frac{1}{2}xy(x+y)e^{-x-y}\mathrm{d}x\mathrm{d}y = 2.$$

It is not equal to $\mathrm{E}(R_1)\mathrm{E}(R_2) = \dfrac{9}{4}$.

★

Example 7.16.

Suppose $X \sim \chi^2(r_1)$ and $Y \sim \chi^2(r_2)$ are two independent random variables. Then the moment generating function of $X + Y$ is given by

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \frac{1}{(1-2t)^{\frac{r_1}{2}}}\frac{1}{(1-2t)^{\frac{r_2}{2}}} = \frac{1}{(1-2t)^{\frac{r_1+r_2}{2}}}, \qquad t < \frac{1}{2},$$

which is the moment generating function of $\chi^2(r_1 + r_2)$. Since moment generating function uniquely determines the distribution, we have

$$X + Y \sim \chi^2(r_1 + r_2).$$

★

7.4 Population Covariance and Correlation

7.4.1 Population Covariance

Definition 7.6.
Let X and Y be random variables with means μ_X and μ_Y, respectively. The *population covariance* between X and Y is defined as

$$\begin{aligned}\sigma_{XY} &= \mathrm{Cov}(X,Y) \\ &= \mathrm{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\ &= \mathrm{E}(XY)-\mu_X\mu_Y.\end{aligned}$$

Note that a more correct notation should be $\sigma_{X,Y}$, but usually by convention σ_{XY} is used as long as it is not confused with the standard deviation of the product XY. ◀

Example 7.17.

In **Example 7.1.**, X is the number of red balls, Y is the number of white balls in a sample of 3 randomly drawn from an urn containing 3 red balls, 4 white balls, and 5 blue balls. The following table shows the joint and marginal pmfs.

Value of X	Value of Y				Total
	0	1	2	3	
0	0.0454	0.1818	0.1364	0.0182	0.3818
1	0.1364	0.2727	0.0818	0	0.4909
2	0.0682	0.0545	0	0	0.1227
3	0.0045	0	0	0	0.0045
Total	0.2545	0.5091	0.2182	0.0182	1.0000

$$\begin{aligned}\mathrm{E}(XY) &= \sum_{x=0}^{3}\sum_{y=0}^{3} xyp(x,y) \\ &= 1\times 1\times 0.2727+1\times 2\times 0.0818+2\times 1\times 0.0545 \\ &= 0.5455, \\ \mathrm{E}(X) &= 1\times 0.4909+2\times 0.1227+3\times 0.0045=0.75, \\ \mathrm{E}(Y) &= 1\times 0.5091+2\times 0.2182+3\times 0.0182=1, \\ \mathrm{Cov}(X,Y) &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)=0.5455-0.75\times 1=-0.2045.\end{aligned}$$

The number of red balls and the number of white balls in the sample are *negatively correlated.*

★

Properties

1. The sign and the magnitude of σ_{XY} reveal the direction and the strength of the linear relationship between X and Y.

   $(X \uparrow, Y \uparrow, \sigma_{XY} > 0;\ X \uparrow, Y \downarrow, \sigma_{XY} < 0)$

2. The magnitude of σ_{XY} depends on the scales of X and Y.

   Let $X' = aX + b$ and $Y' = cY + d$ where a and c are non-zero constants. Then,

$$\begin{aligned}
\mathrm{Cov}(X', Y') &= \mathrm{Cov}(aX + b, cY + d) \\
&= \mathrm{E}\left\{[aX + b - \mathrm{E}(aX + b)]\,[cY + d - \mathrm{E}(cY + d)]\right\} \\
&= \mathrm{E}\left[(aX - a\mu_X)(cY - c\mu_Y)\right] \\
&= ac\mathrm{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= ac\mathrm{Cov}(X, Y).
\end{aligned}$$

3. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

4. $\mathrm{Cov}(X, c) = 0$ for any constant c.

5. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

6. $\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathrm{Cov}(X_i, Y_j)$.

7. $$\begin{aligned}
\mathrm{Var}(X + Y) &= \mathrm{Cov}(X + Y, X + Y) \\
&= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X) + \mathrm{Cov}(Y, Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y).
\end{aligned}$$

   In general,

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

Example 7.18.

Suppose that a couple is drawn at random from a large population of working couples. In thousand dollar units, let

$$X = \text{man's income}, \qquad Y = \text{woman's income}.$$

Then the couple's total income is the sum

$$S = X + Y.$$

Suppose their pension contribution is 10% of the man's income, and 20% of the woman's income. Then the couple's total pension contribution is a weighted sum:

$$W = 0.1X + 0.2Y.$$

Suppose we know that the average man's income is $\mathrm{E}(X) = 20$ and the average woman's is $\mathrm{E}(Y) = 16$, then the average total income is

$$\mathrm{E}(S) = \mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y) = 20 + 16 = 36,$$

and the average total pension contribution is

$$\mathrm{E}(W) = \mathrm{E}(0.1X + 0.2Y) = 0.1\mathrm{E}(X) + 0.2\mathrm{E}(Y) = 2 + 3.2 = 5.2.$$

Furthermore, if we were given the joint distribution of X and Y from which we calculated $\sigma_X^2 = 60$, $\sigma_Y^2 = 70$ and $\sigma_{XY} = 49$, then

$$\begin{aligned}
\mathrm{Var}(S) &= \mathrm{Var}(X + Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) \\
&= 60 + 70 + 2 \times 49 \\
&= 228, \\
\sigma_S &= \sqrt{228} = 15.1,
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(W) &= \mathrm{Var}(0.1X + 0.2Y) \\
&= (0.1)^2\mathrm{Var}(X) + (0.2)^2\mathrm{Var}(Y) + 2(0.1)(0.2)\mathrm{Cov}(X, Y) \\
&= (0.01)(60) + (0.04)(70) + 2(0.02)(49) \\
&= 5.36, \\
\sigma_W &= \sqrt{5.36} = 2.32.
\end{aligned}$$

★

Example 7.19. (Markowitz's portfolio selection model)

Suppose an investment portfolio involves two assets with random returns denoted by R_1 and R_2, according to the weighting w and $1-w$ respectively ($0 < w < 1$). The portfolio return can be expressed as $R = wR_1 + (1-w)R_2$, with expected value and variance given by:

$$\begin{aligned} \text{Expected return:} \quad & \mu_R = w\mu_{R_1} + (1-w)\mu_{R_2} \\ \text{Return variance:} \quad & \sigma_R^2 = w^2\sigma_{R_1}^2 + (1-w)^2\sigma_{R_2}^2 + 2w(1-w)\sigma_{R_1R_2} \end{aligned}$$

From the first equation we can express w in terms of μ_{R_1}, μ_{R_2} and μ_R as $w = \dfrac{\mu_R - \mu_{R_2}}{\mu_{R_1} - \mu_{R_2}}$.

Substituting in the second equation yields

$$\sigma_R^2 = \frac{(\mu_R - \mu_{R_2})^2\sigma_{R_1}^2 + (\mu_R - \mu_{R_1})^2\sigma_{R_2}^2 - 2(\mu_R - \mu_{R_2})(\mu_R - \mu_{R_1})\sigma_{R_1R_2}}{(\mu_{R_1} - \mu_{R_2})^2}.$$

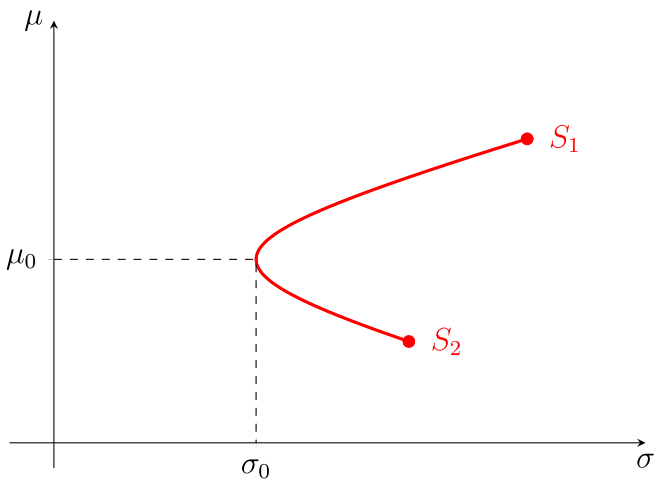
Therefore the return variance σ_R^2 that measures the risk of the portfolio can be expressed as a quadratic function of the expected return μ_R. By completing square, we have

$$\sigma_R^2 = A(\mu_R - \mu_0)^2 + \sigma_0^2,$$

where

$$\begin{aligned} \mu_0 &= \frac{\mu_{R_2}\sigma_{R_1}^2 + \mu_{R_1}\sigma_{R_2}^2 - (\mu_{R_1} + \mu_{R_2})\sigma_{R_1R_2}}{\sigma_{R_1}^2 + \sigma_{R_2}^2 - 2\sigma_{R_1R_2}}, \\ A &= \frac{\sigma_{R_1}^2 + \sigma_{R_2}^2 - 2\sigma_{R_1R_2}}{(\mu_{R_1} - \mu_{R_2})^2} > 0, \\ \sigma_0^2 &= \frac{\sigma_{R_1}^2\sigma_{R_2}^2 - \sigma_{R_1R_2}^2}{\sigma_{R_1}^2 + \sigma_{R_2}^2 - 2\sigma_{R_1R_2}} > 0. \end{aligned}$$

The following graph shows how the expected return changes with respect to the risk (volatility, the standard deviation of the return):



Each point on the curve represents the outcome of a particular portfolio. For each fixed value of the risk, there can be two corresponding portfolios, one with lower expected return and one with higher expected return. The portfolios located at the upper branch of the curve provides the greatest expected return for a given level of risk, thereby are called the *efficient portfolios*. For the portfolios in such region, the greater the risk come the greater reward.

A risk-averse investor may choose the portfolio such that the risk is minimized. It can be achieved by setting the weighting w in such a way that $\mu_R = \mu_0$, i.e.,

$$w = \frac{\mu_0 - \mu_{R_2}}{\mu_{R_1} - \mu_{R_2}} = \frac{\sigma_{R_2}^2 - \sigma_{R_1 R_2}}{\sigma_{R_1}^2 + \sigma_{R_2}^2 - 2\sigma_{R_1 R_2}},$$

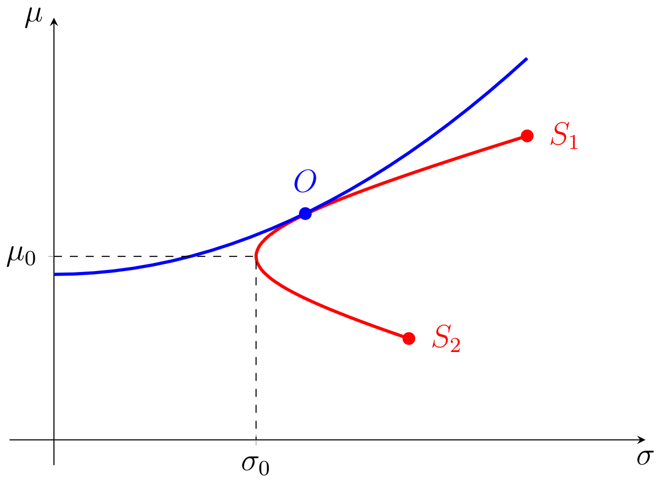
with minimum variance

$$\sigma_0^2 = \frac{\sigma_{R_1}^2 \sigma_{R_2}^2 - \sigma_{R_1 R_2}^2}{\sigma_{R_1}^2 + \sigma_{R_2}^2 - 2\sigma_{R_1 R_2}}.$$

For an investor who is not totally risk-averse, we may define a *utility function* which increases with respect to increased μ_R, and decreases with respect to increased σ_R, so that a portfolio with high risk will be penalized. For example, the following utility function is used:

$$U = \mu_R - k\sigma_R^2,$$

where $k > 0$ represents a measure of the level of risk aversion of the investor. According to this utility function, an optimal portfolio can be determined by maximizing U subject to the constraint $\sigma_R^2 = A(\mu_R - \mu_0)^2 + \sigma_0^2$. It can be done easily by the *method of Lagrange multipliers*, or by *nonlinear programming* as shown in the graph below:



The upper curve is a curve that takes the form of $U_0 = \mu_R - k\sigma_R^2$ for some U_0 and tangent to the quadratic curve representing the constraint. The optimal portfolio is the one corresponding to point O at the point of tangent.

Example 7.20. (Sampling without replacement from a finite population)

Suppose we randomly draw n balls from an urn with m red balls and $N-m$ white balls. Let X be the number of red balls in our sample. Then X has a hypergeometric distribution with pmf

$$p(x) = \Pr(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \qquad \max\left[n-(N-m), 0\right] \le x \le \min(n, m).$$

Direct derivation of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ may be difficult.

Let

$$Y_i = \begin{cases} 1, & \text{if the } i\text{-th ball drawn is red;} \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$X = \sum_{i=1}^{n} Y_i.$$

Consider

$$\begin{aligned} \mathrm{E}(Y_i) &= \mathrm{E}(Y_i^2) = \Pr(Y_i = 1) = \frac{m}{N}, \\ \mathrm{Var}(Y_i) &= \frac{m}{N} - \frac{m^2}{N^2} = \frac{m(N-m)}{N^2}. \end{aligned}$$

For $i \ne j$,

$$\begin{aligned} \mathrm{E}(Y_iY_j) &= \Pr(Y_i = 1, Y_j = 1) = \frac{m}{N}\frac{m-1}{N-1}, \\ \mathrm{Cov}(Y_i, Y_j) &= \frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2} = -\frac{m(N-m)}{N^2(N-1)}. \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathrm{E}(Y_i) = \sum_{i=1}^{n} \frac{m}{N} = \frac{nm}{N}, \\ \mathrm{Var}(X) &= \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) \\ &= \sum_{i=1}^{n} \mathrm{Var}(Y_i) + 2\sum_{i<j} \mathrm{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^{n} \frac{m(N-m)}{N^2} + 2\sum_{i<j}\left[-\frac{m(N-m)}{N^2(N-1)}\right] \\ &= \frac{nm(N-m)}{N^2} - 2\frac{n(n-1)}{2}\frac{m(N-m)}{N^2(N-1)} \\ &= \left(\frac{N-n}{N-1}\right) n\frac{m}{N}\left(1 - \frac{m}{N}\right). \end{aligned}$$

★

7.4.2 Coefficient of Correlation

Since σ_{XY} depends on the scales of X and Y, it is difficult to determine the strength of the linear relationship between X and Y. Thus we need a standardized measure which is invariant under linear transformation of X and Y.

Definition 7.7.
Let X and Y be random variables with covariance σ_{XY}, standard deviations $\sigma_X = \sqrt{\mathrm{Var}(X)}$ and $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$. The *population correlation coefficient* between X and Y is defined as

$$\rho = \rho_{XY} = \mathrm{Corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Similar to covariance, a more correct notation should be $\rho_{X,Y}$. ◀

Example 7.21.
In **Example 7.17.**, $\sigma_{XY} = -0.2045$, $\sigma_X = 0.6784$, $\sigma_Y = 0.7385$.
Hence,

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-0.2045}{0.6784 \times 0.7385} = -0.4082.$$

The number of red balls and the number of white balls in the sample are slightly negatively correlated.

★

Theorem 7.2. (Cauchy-Schwarz Inequality)
Let X and Y be random variables with finite second moments. Then

$$[\mathrm{E}(XY)]^2 \le \mathrm{E}(X^2)\mathrm{E}(Y^2).$$

The equality holds if and only if either $\Pr(X=0)=1$ or $\Pr(Y=aX)=1$ for some constant a, i.e., X and Y are proportional. ▲

Proof. For any $t \in \mathbb{R}$ and random variables X and Y, consider

$$0 \le \mathrm{E}\left[(tX+Y)^2\right] = t^2\mathrm{E}(X^2) + 2t\mathrm{E}(XY) + \mathrm{E}(Y^2).$$

Put $t = -\dfrac{\mathrm{E}(XY)}{\mathrm{E}(X^2)}$,

$$\begin{aligned} 0 &\le \frac{[\mathrm{E}(XY)]^2}{\mathrm{E}(X^2)} - \frac{2\,[\mathrm{E}(XY)]^2}{\mathrm{E}(X^2)} + \mathrm{E}(Y^2) \\ \frac{[\mathrm{E}(XY)]^2}{\mathrm{E}(X^2)} &\le \mathrm{E}(Y^2) \\ [\mathrm{E}(XY)]^2 &\le \mathrm{E}(X^2)\mathrm{E}(Y^2). \end{aligned}$$

□

Properties of the correlation coefficient

1. $-1 \leq \rho \leq 1$

   Proof. By Cauchy-Schwarz inequality,

$$\begin{aligned} \sigma_{XY}^2 &= \{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]\}^2 \\ &\leq \mathrm{E}\left[(X-\mu_X)^2\right]\mathrm{E}\left[(Y-\mu_Y)^2\right] = \sigma_X^2\sigma_Y^2. \end{aligned}$$

   That is,

$$\rho^2 = \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1.$$

   The equality holds ($\rho = \pm 1$) when $X - \mu_X$ and $Y - \mu_Y$ are proportional, i.e., when

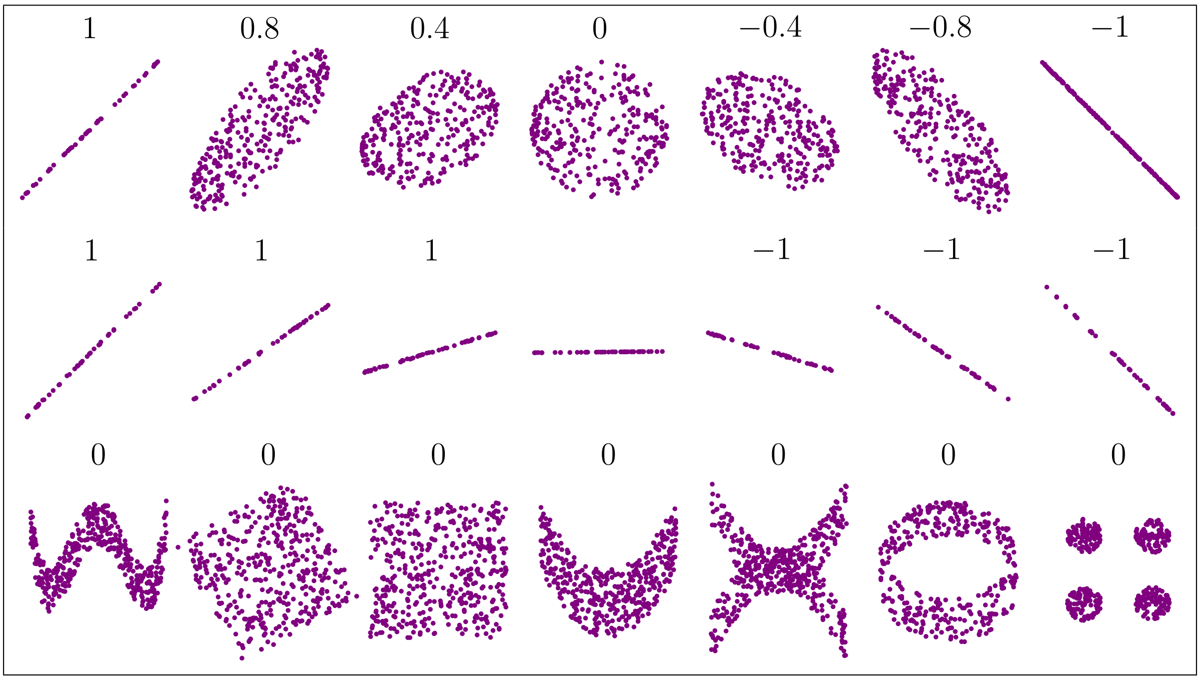
$$\Pr(Y - \mu_Y = a(X - \mu_X)) = 1,$$

   for some constant a, or

$$\Pr(Y = aX + b) = 1,$$

   for some constant b, that is, X and Y are perfectly linearly related. $\square$

   Note that when X or Y (or both) is a constant, ρ is undefined.



2. ρ is invariant under linear transformation of X and Y.

   Proof. Let $X' = aX + b$ and $Y' = cY + d$ where a and c are non-zero constants. Then,

$$\begin{aligned}\text{Corr}(X', Y') &= \frac{\text{Cov}(aX+b, cY+d)}{\sqrt{\text{Var}(aX+b)\text{Var}(cY+d)}} \\ &= \frac{\text{Cov}(aX, cY)}{\sqrt{\text{Var}(aX)\text{Var}(cY)}} \\ &= \frac{ac\text{Cov}(X,Y)}{\sqrt{a^2\text{Var}(X)c^2\text{Var}(Y)}} \\ &= \frac{ac\text{Cov}(X,Y)}{|ac|\sqrt{\text{Var}(X)\text{Var}(Y)}} \\ &= \text{sgn}(ac)\text{Corr}(X,Y)\end{aligned}$$

□

3. If X and Y are independent, then

$$\text{Cov}(X,Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y) = \text{E}(X)\text{E}(Y) - \text{E}(X)\text{E}(Y) = 0,$$

   and hence $\rho = 0$ if X and Y are not constants.

4. If X and Y are independent, then

$$\text{Var}(X+Y) = \text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y).$$

Remarks

1. The converse of property 3 need not be true. That is, $\rho = 0$ does not imply X and Y are independent.
2. Correlation coefficient measures the strength of the linear relationship only. It may be possible that X and Y are strongly related but that the relation is curvilinear, and ρ would be nearly zero.
3. An observed correlation may be due to a third unknown casual variable.

Example 7.22. (Random point at the circumference of a unit circle)

Consider

$$\Theta \sim \mathrm{U}(0, 2\pi), \qquad X = \cos\Theta, \qquad Y = \sin\Theta.$$

Obviously X and Y are not independent because $\Pr(X^2 + Y^2 = 1) = 1$.

However,

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^{2\pi} \frac{\cos\theta}{2\pi} \mathrm{d}\theta = \left[\frac{\sin\theta}{2\pi}\right]_0^{2\pi} = 0, \\
\mathrm{E}(Y) &= \int_0^{2\pi} \frac{\sin\theta}{2\pi} \mathrm{d}\theta = -\left[\frac{\cos\theta}{2\pi}\right]_0^{2\pi} = 0, \\
\mathrm{E}(XY) &= \int_0^{2\pi} \frac{\cos\theta\sin\theta}{2\pi} \mathrm{d}\theta = \int_0^{2\pi} \frac{2\sin\theta\cos\theta}{4\pi} \mathrm{d}\theta = \int_0^{2\pi} \frac{\sin 2\theta}{4\pi} \mathrm{d}\theta = -\left[\frac{\cos 2\theta}{8\pi}\right]_0^{2\pi} = 0.
\end{aligned}$$

Hence, $\mathrm{Cov}(X, Y) = 0$, i.e., X and Y are uncorrelated.

★

∼ End of Chapter 7 ∼