Multivariate distributions and measures of dependence between random variables *

Jean-Marie Dufour †
McGill University

First version: September 1980
Revised: October 1995, July 2011, September 2011, September 2017
This version: September 2017
Compiled: March 24, 2019, 10:33

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca`. Web page: `http://www.jeanmariedufour.com`

Contents

1. Random variables

1.1 In general, economic theory specifies exact relations between economic variables. Even a superficial examination of economic data indicates it is not (almost never) possible to find such relationships in actual data. Instead, we have relations of the form:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

where ε_t can be interpreted as a "random variable".

1.2 Definition *A random variable* (r.v.) X *is a variable whose behavior can be described by a "probability law". If X takes its values in the real numbers, the probability law of X can be described by a "distribution function":*

$$F_X(x) = \mathsf{P}\left[X \leq x\right]$$

1.3 If X is continuous, there is a "density function" $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx\,.$$

The mean and variance of X are given by:

$$\mu_X = \mathsf{E}(X) = \int_{-\infty}^{+\infty} x\, dF_X(x) \qquad \text{(general case)}$$

$$= \int_{-\infty}^{+\infty} x\, f_X(x)\; dx \qquad \text{(continuous case)}$$

$$
\begin{aligned}
\mathsf{V}(X) = \sigma_X^2 = \mathsf{E}\left[(X-\mu_X)^2\right] &= \int_{-\infty}^{+\infty} (x-\mu_X)^2 \, dF_X(x) && \text{(general case)} \\
&= \int_{-\infty}^{+\infty} (x-\mu_X)^2 F_X(x) \, dx && \text{(continuous case)} \\
&= \mathsf{E}\left(X^2\right) - [\mathsf{E}(X)]^2
\end{aligned}
$$

1.4 It is easy to characterize relations between two non-random variables x and y :

$$g(x, y) = 0$$

or (in certain cases)

$$y = f(x) \, .$$

How does one characterize the links or relations between random variables? The behavior of a pair $(X, Y)'$ is described by a joint distribution function:

$$
\begin{aligned}
F(x,y) &= \mathsf{P}[X \leq x, \, Y \leq y] \\
&= \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) \, dx \, dy && \text{(continuous case.)}
\end{aligned}
$$

We call $f(x, y)$ the joint density function of $(X, Y)'$. More generally, if we consider k r.v.'s $X_1, X_2, \ldots, X_k$, their behavior can be described through a k-dimensional distribution function:

$$F(x_1, x_2, \ldots, x_k) = \mathsf{P}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k]$$

$$= \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(x_1, x_2, \ldots, x_k)\, dx_1 dx_2 \cdots dx_k \qquad \text{(continuous case)}$$

where $f(x_1, x_2, \ldots, x_k)$ is the joint density function of $X_1, X_2, \ldots, X_k$.

2. Covariances and correlations

We often wish to have a simple measure of association between two random variables X and Y. The notions of "covariance" and "correlation" provide such measures of association. Let X and Y be two *r.v.*'s with means μ_X and μ_Y and finite variances σ_X^2 and σ_Y^2. Below *a.s.* means "almost surely" (with probability 1).

2.1 Definition *The covariance between X and Y is defined by*

$$\mathsf{C}(X,Y) \equiv \sigma_{XY} \equiv \mathsf{E}\left[(X-\mu_X)(Y-\mu_Y)\right] .$$

2.2 Definition *Suppose $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$. Then the correlation between X and Y is defined by*

$$\rho(X,Y) \equiv \rho_{XY} \equiv \sigma_{XY}/\sigma_X\sigma_Y .$$

When $\sigma_X^2 = 0$ or $\sigma_Y^2 = 0$, we set $\rho_{XY} = 0$.

2.3 Theorem *The covariance and correlation between X and Y satisfy the following properties:*

(a) $\sigma_{XY} = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$;

(b) $\sigma_{XY} = \sigma_{YX}$, $\rho_{XY} = \rho_{YX}$;

(c) $\sigma_{XX} = \sigma_X^2$, $\rho_{XX} = 1$;

(d) $\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2$; (*Cauchy-Schwarz inequality*)

(e) $-1 \leq \rho_{XY} \leq 1$;

(f) *X and Y are independent* $\Rightarrow \sigma_{XY} = 0 \Rightarrow \rho_{XY} = 0$;

(g) *if* $\sigma_X^2 \neq 0$ *and* $\sigma_Y^2 \neq 0$,

$$\rho_{XY}^2 = 1 \Leftrightarrow \left[\exists \text{ two constants } a \text{ and } b \text{ such that } a \neq 0 \text{ and } Y = aX + b \text{ a.s.}\right]$$

PROOF (a)

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

(b) et (c) are immediate. To get (d), we observe that

$$\begin{aligned}
\mathsf{E}\left\{\left[Y-\mu_Y-\lambda(X-\mu_X)\right]^2\right\} &= \mathsf{E}\left\{\left[(Y-\mu_Y)-\lambda(X-\mu_X)\right]^2\right\} \\
&= \mathsf{E}\left\{(Y-\mu_Y)^2 - 2\lambda(X-\mu_X)(Y-\mu_Y) + \lambda^2(X-\mu_X)^2\right\} \\
&= \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 \geq 0 \ .
\end{aligned}$$

for any arbitrary constant λ. In other words, the second-order polynomial $g(\lambda) = \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2$ cannot take negative values. This can happen only if the equation

$$\lambda^2\sigma_X^2 - 2\lambda\sigma_{XY} + \sigma_Y^2 = 0 \tag{2.1}$$

does not have two distinct real roots, i.e. the roots are either complex or identical. The roots of equation (2.1). are given by

$$\lambda = \frac{2\sigma_{XY} \pm \sqrt{4\sigma_{XY}^2 - 4\sigma_X^2\sigma_Y^2}}{2\sigma_X^2} = \frac{\sigma_{XY} \pm \sqrt{\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2}}{\sigma_X^2} .$$

Distinct real roots are excluded when $\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 \leq 0$, hence

$$\sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 .$$

(e)

$$\begin{aligned} \sigma_{XY}^2 \leq \sigma_X^2\sigma_Y^2 &\Rightarrow -\sigma_X\sigma_Y \leq \sigma_{XY} \leq \sigma_X\sigma_Y \\ &\Rightarrow -1 \leq \rho_{XY} \leq 1 . \end{aligned}$$

(f)

$$\begin{aligned} \sigma_{XY} &= \mathsf{E}\{(X-\mu_X)(Y-\mu_Y)\} = \mathsf{E}(X-\mu_X)\,\mathsf{E}(Y-\mu_Y) \\ &= [\mathsf{E}(X)-\mu_X]\,[\mathsf{E}(Y)-\mu_Y] = 0, \\ \rho_{XY} &= \sigma_{XY} \big/ \sigma_X\sigma_Y = 0. \end{aligned}$$

Note the reverse implication does not hold in general, *i.e.*,

$$\rho_{XY} = 0 \neq> X \text{ and } Y \text{ are independent}$$

(g) 1) Necessity of the condition. If $Y = aX + b$, then

$$\mathsf{E}(Y) = a\mathsf{E}(X) + b = a\mu_X + b\ ,\ \ \sigma_Y^2 = a^2\sigma_X^2\ ,$$

and

$$\sigma_{XY} = \mathsf{E}\left[(Y - \mu_Y)(X - \mu_X)\right] = \mathsf{E}\left[a(X - \mu_X)(X - \mu_X)\right] = a\sigma_X^2\ .$$

Consequently,

$$\rho_{XY}^2 = \frac{a^2\sigma_X^4}{a^2\sigma_X^2\sigma_X^2} = 1\ .$$

2) Sufficiency of the condition. If $\rho_{XY}^2 = 1$, then

$$\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 = 0.$$

In this case, the equation

$$\mathsf{E}\left\{\left[(Y - \mu_Y) - \lambda(X - \mu_X)\right]^2\right\} = \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 = 0$$

has one and only one root

$$\lambda = \frac{2\sigma_{XY}}{2\sigma_X^2} = \sigma_{XY}/\sigma_X^2\ ,$$

so that

$$\mathsf{E}\left\{\left[\left(Y\sigma_Y^2 - \mu_Y\right) - \frac{\sigma_{XY}}{\sigma_X^2}(X - \mu_X)\right]^2\right\} = 0$$

and

$$\mathsf{P}\left[(Y-\mu_Y)-\frac{\sigma_{XY}}{\sigma_X^2}(X-\mu_X)=0\right]=\mathsf{P}\left[Y=\frac{\sigma_{XY}}{\sigma_X^2}X+\left(\mu_Y-\frac{\sigma_{XY}}{\sigma_X^2}\mu_X\right)\right]=1$$

We can thus write:

$$Y=aX+b \text{ with probability 1}$$

where $a=\sigma_{XY}/\sigma_X^2$ and $b=\mu_Y-\frac{\sigma_{XY}}{\sigma_y^2}\mu_X$. $\square$

3. Alternative interpretations of covariances and correlations

Highly correlated random variables tend to be "close". This feature can be explicated in different ways:

1. by looking at the distribution of the difference $Y-X$;
2. by looking at the difference of two variances (polarization identity);
3. by looking at the linear regression of Y on X;
4. through a "decoupling" representation of covariances and correlations.

3.1. Difference between two correlated random variables

First, we can look at the difference of two random variables X and Y. It is easy to see that

$$\begin{aligned}
E[(Y-X)^2] &= E\left\{\left([(Y-\mu_Y)-(X-\mu_X)]-(\mu_Y-\mu_X)\right)^2\right\} \\
&= E\left\{\left([(Y-\mu_Y)-(X-\mu_X)]\right)^2\right\}+(\mu_Y-\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2-2\sigma_{XY}+(\mu_Y-\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2-2\rho_{XY}\sigma_X\sigma_Y+(\mu_Y-\mu_X)^2\,. \qquad (3.1)
\end{aligned}$$

$E[(Y-X)^2]$ has three components: (1) a *variance component* $\sigma_Y^2+\sigma_X^2$; (2) a *covariance component* $-2\sigma_{XY}$; (3) a *mean component* $(\mu_Y-\mu_X)^2$. Equation (3.1) shows clearly that $E[(Y-X)^2]$ tends to be large, when they have very different means or variances.

Since $|\rho_{XY}| \leq 1$, it is interesting to observe that

$$(\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2 \leq E[(Y-X)^2] \leq (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2\,, \tag{3.2}$$

and

$$E[(Y-X)^2] \leq \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2 \leq (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2\,,\ \text{if } \rho_{XY} \geq 0\,, \tag{3.3}$$

$$E[(Y-X)^2] \geq \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2 \geq (\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2\,,\ \text{if } \rho_{XY} \leq 0\,, \tag{3.4}$$

$$E[(Y-X)^2] = \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2\,,\ \text{if } \rho_{XY} = 0\,. \tag{3.5}$$

$E[(Y-X)^2]$ reaches its minimum value when $\rho_{XY} = 1$, and its maximal value when $\rho_{XY} = -1$:

$$E[(Y-X)^2] = (\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2\,, \quad \text{if } \rho_{XY} = 1\,, \tag{3.6}$$

$$E[(Y-X)^2] = (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2\,, \quad \text{if } \rho_{XY} = -1\,. \tag{3.7}$$

If $\sigma_Y^2 > 0$, we can also write:

$$\left(1 - \frac{\sigma_X}{\sigma_Y}\right)^2 + \left(\frac{\mu_Y - \mu_X}{\sigma_Y}\right)^2 \leq \frac{E[(Y-X)^2]}{\sigma_Y^2} \leq \left(1 + \frac{\sigma_X}{\sigma_Y}\right)^2 + \left(\frac{\mu_Y - \mu_X}{\sigma_Y}\right)^2\,. \tag{3.8}$$

The inequalities (3.2) - (3.5) also entail similar properties for $X+Y$:

$$(\sigma_X - \sigma_Y)^2 + (\mu_X + \mu_Y)^2 \leq E[(X+Y)^2] \leq (\sigma_X + \sigma_Y)^2 + (\mu_X + \mu_Y)^2\,, \tag{3.9}$$

$$E[(X+Y)^2] \leq \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \leq (\sigma_Y + \sigma_X)^2 + (\mu_X + \mu_Y)^2\,,\ \text{if } \rho_{XY} \leq 0\,, \tag{3.10}$$

$$E[(X+Y)^2] \geq \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \geq (\sigma_X - \sigma_Y)^2 + (\mu_X + \mu_Y)^2\,,\ \text{if } \rho_{XY} \geq 0\,, \tag{3.11}$$

$$E[(Y+X)^2] = \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2\,,\ \text{if}\ \rho_{XY} = 0\,. \tag{3.12}$$

From (3.1), it is also easy to see that

$$E\left[\left(\frac{Y}{\sigma_Y} - \frac{X}{\sigma_X}\right)^2\right] = 2(1-\rho_{XY}) + \left(\frac{\mu_Y}{\sigma_Y} - \frac{\mu_X}{\sigma_X}\right)^2\,. \tag{3.13}$$

Let

$$\tilde{X} = \frac{X-\mu_X}{\sigma_X}\,, \quad \tilde{Y} = \frac{Y-\mu_Y}{\sigma_Y}\,, \quad \rho\left(\tilde{X},\tilde{Y}\right) = \rho\left(X,Y\right) = \rho_{XY}\,, \tag{3.14}$$

where we set $\tilde{X} = 0$ if $\sigma_X = 0$, and $\tilde{Y} = 0$ if $\sigma_Y = 0$. We then have:

$$E(\tilde{X}) = E(\tilde{Y}) = 0\,, \quad \mathsf{V}(\tilde{X}) = \mathsf{V}(\tilde{Y}) = 1\,, \tag{3.15}$$

and

$$E[(\tilde{Y}-\tilde{X})^2] = 2(1-\rho_{XY})\,. \tag{3.16}$$

Since

$$X = \mu_X + \sigma_X\tilde{X}\,, \quad Y = \mu_Y + \sigma_Y\tilde{Y}\,, \tag{3.17}$$

we get

$$\begin{aligned} E[(Y-X)^2] &= E\left\{[(\mu_Y + \sigma_Y\tilde{Y}) - (\mu_X + \sigma_X\tilde{X})]^2\right\} \\ &= E\left\{[(\sigma_Y\tilde{Y} - \sigma_X\tilde{X}) + (\mu_Y - \mu_X)]^2\right\} \\ &= E\left\{[(\sigma_Y\tilde{Y} - \sigma_X\tilde{X}) + (\mu_Y - \mu_X)]^2\right\} \end{aligned}$$

$$= E[(\sigma_Y \tilde{Y} - \sigma_X \tilde{X})^2] + (\mu_Y - \mu_X)^2 \tag{3.18}$$

hence

$$\begin{aligned} E[(Y-X)^2] &= \sigma_Y^2 E\left[\left(\tilde{Y} - \frac{\sigma_X}{\sigma_Y}\tilde{X}\right)^2\right] + (\mu_Y - \mu_X)^2 \\ &= \sigma_Y^2 \left[1 + \left(\frac{\sigma_X}{\sigma_Y}\right)^2 - 2\left(\frac{\sigma_X}{\sigma_Y}\right)\rho_{XY}\right] + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y \neq 0, \end{aligned} \tag{3.19}$$

and

$$E[(Y-X)^2] = \sigma_X^2 + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y = 0. \tag{3.20}$$

If the variances of X and Y are the same, i.e.

$$\sigma_Y^2 = \sigma_X^2, \tag{3.21}$$

we have:

$$\begin{aligned} E[(Y-X)^2] &= 2\sigma_Y^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2 \\ &= 2\sigma_X^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2. \end{aligned} \tag{3.22}$$

If the means and variances of X and Y are the same, i.e.

$$\mu_Y = \mu_X \text{ and } \sigma_Y^2 = \sigma_X^2, \tag{3.23}$$

we have:

$$E[(Y-X)^2] = 2\sigma_Y^2\,(1-\rho_{XY}) = 2\sigma_X^2\,(1-\rho_{XY}) \tag{3.24}$$

and

$$0 \le E[(Y-X)^2] \le 4\sigma_X^2 \tag{3.25}$$

so that

$$E[(Y-X)^2] = 0 \text{ and } \mathsf{P}[Y=X] = 1, \text{ if } \rho_{XY} = 1\,, \tag{3.26}$$

and, using Chebyshev's inequality,

$$\mathsf{P}[|Y-X|>c] \le \frac{E[(Y-X)^2]}{c^2} = \frac{2\sigma_X^2\,(1-\rho_{XY})}{c^2} \text{ for any } c>0\,, \tag{3.27}$$

$$\mathsf{P}\,[|Y-X|>c\sigma_X] \le \frac{E[(Y-X)^2]}{\sigma_X^2\,c^2} = \frac{2(1-\rho_{XY})}{c^2} \text{ for any } c>0\,. \tag{3.28}$$

If $\mu_Y = \mu_X$ and $\sigma_Y^2 = \sigma_X^2 > 0$, we also have:

$$E[(Y-X)^2] = 0 \Leftrightarrow \rho_{XY} = 1\,, \tag{3.29}$$

$$E[(Y-X)^2] = 2\sigma_X^2 \Leftrightarrow \rho_{XY} = 0\,, \tag{3.30}$$

$$E[(Y-X)^2] = 4\sigma_X^2 \Leftrightarrow \rho_{XY} = -1\,. \tag{3.31}$$

Since

$$\sigma_Y(\tilde{Y}-\tilde{X}) = Y-\mu_Y - \frac{\sigma_Y}{\sigma_X}(X-\mu_X) = Y - \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) - \frac{\sigma_Y}{\sigma_X}X\,, \tag{3.32}$$

the linear function

$$L_0(X) = \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) + \frac{\sigma_Y}{\sigma_X}X \tag{3.33}$$

can be viewed as a "forecast" of Y based on X such that

$$E[(Y - L_0(X))^2] = \sigma_Y^2 E[(\tilde{Y} - \tilde{X})^2] = 2\sigma_Y^2(1 - \rho_{XY})\,. \tag{3.34}$$

It is then of interest to note that

$$E[(Y - L_0(X))^2] \leq E[(Y - \mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} \geq 0.5\,, \tag{3.35}$$

with

$$E[(Y - L_0(X))^2] < E[(Y - \mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} > 0.5 \tag{3.36}$$

when $\sigma_Y^2 > 0$. Thus $L_0(X)$ provides a "better forecast" of Y than the mean of Y, when $\rho_{XY} > 0.5$. If $\rho_{XY} < 0.5$ and $\sigma_Y^2 > 0$, the opposite holds: $E[(Y - L_0(X))^2] > \sigma_Y^2$.

3.2. Polarization identity

Since

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y), \tag{3.37}$$

$$V(X-Y) = V(X) + V(Y) - 2C(X,Y), \tag{3.38}$$

it is easy to see that

$$C(X,Y) = \frac{1}{4}[V(X+Y) - V(X-Y)]. \tag{3.39}$$

(3.39) is sometimes called the "polarization identity". Further,

$$\rho(X,Y) = \frac{1}{4}\frac{V(X+Y) - V(X-Y)}{\sigma_X\,\sigma_Y} = \frac{1}{4}\left[\frac{\sigma^2_{X+Y}}{\sigma_X\,\sigma_Y} - \frac{\sigma^2_{X-Y}}{\sigma_X\,\sigma_Y}\right] \tag{3.40}$$

and, if $V(X) = V(Y) = 1$,

$$\rho(X,Y) = \frac{V(X+Y) - V(X-Y)}{4} = \frac{\sigma^2_{X+Y} - \sigma^2_{X-Y}}{4}. \tag{3.41}$$

On $X+Y$ and $X-Y$, it also interesting to observe that

$$C(X+Y, X-Y) = [V(X) - V(Y)] + [C(Y,X) - C(X,Y)] = V(X) - V(Y) \tag{3.42}$$

so

$$C\big((X+Y)/2, X-Y\big) = C(X+Y, X-Y) = 0, \quad \text{if } V(X) = V(Y). \tag{3.43}$$

This holds irrespective of the covariance between between X and Y. In particular, if the vector (X,Y) is multinormal $X+Y$ and $X-Y$ are independent when $V(X)=V(Y)$.

4. Covariance matrices

Consider now k r.v.´s $X_1, X_2, \ldots, X_k$ such that

$$\begin{aligned} \mathsf{E}(X_i) &= \mu_i \, , \; i = 1, \ldots, k \, , \\ \mathsf{C}(X_i, X_j) &= \sigma_{ij} \, , \; i, j = 1, \ldots, k \, . \end{aligned}$$

We often wish to compute the mean and variance of a linear combination of $X_1, \ldots, X_k$:

$$\Sigma_{i=1}^{k} a_i X_i = a_1 X_1 + a_2 X_2 + \cdots + a_k X_k \, .$$

It is easily verified that

$$\mathsf{E}\left[\Sigma_{i=1}^{k} a_i X_i\right] = \Sigma_{i=1}^{k} a_i \mu_i$$

and

$$\begin{aligned} \mathsf{V}\left[\Sigma_{i=1}^{k} a_i X_i\right] &= \mathsf{E}\left\{\left[\Sigma_{i=1}^{k} a_i (X_i - \mu_i)\right]\left[\Sigma_{j=1}^{k} a_j \left(X_j - \mu_j\right)\right]\right\} \\ &= \Sigma_{i=1}^{k} \Sigma_{j=1}^{k} a_i a_j \sigma_{ij} \, . \end{aligned}$$

Since such formulae may often become cumbersome, it will be convenient to use vector and matrix notation
We define a random vector $\mathbf{X}$ and its mean value $\mathsf{E}(\mathbf{X})$ by:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \; , \; \mathsf{E}(\mathbf{X}) = \begin{pmatrix} \mathsf{E}(X_1) \\ \vdots \\ \mathsf{E}(X_k) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} \equiv \mu_X \, .$$

Similarly, we define a random matrix M and its mean value $\mathsf{E}(M)$ by:

$$M = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \ldots & X_{mn} \end{bmatrix}, \ \mathsf{E}(M) = \begin{bmatrix} \mathsf{E}(X_{11}) & \mathsf{E}(X_{12}) & \ldots & \mathsf{E}(X_{1n}) \\ \mathsf{E}(X_{21}) & \mathsf{E}(X_{22}) & \ldots & \mathsf{E}(X_{2n}) \\ \vdots & \vdots & & \vdots \\ \mathsf{E}(X_{m1}) & \mathsf{E}(X_{m2}) & \ldots & \mathsf{E}(X_{mn}) \end{bmatrix}$$

where the X_{ij} are *r.v.´*s. To a random vector $\mathbf{X}$, we can associate a covariance matrix $\mathsf{V}(\mathbf{X})$:

$$\begin{aligned} \mathsf{V}(\mathbf{X}) &= \mathsf{E}\left\{[\mathbf{X} - \mathsf{E}(\mathbf{X})][\mathbf{X} - \mathsf{E}(\mathbf{X})]'\right\} = \mathsf{E}\left\{[\mathbf{X} - \mu_X][\mathbf{X} - \mu_X]'\right\} \\ &= \mathsf{E}\left\{ \begin{bmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & (X_1 - \mu_1)(X_2 - \mu_2) & \ldots & (X_1 - \mu_1)(X_k - \mu_k) \\ \vdots & \vdots & & \vdots \\ (X_k - \mu_k)(X_1 - \mu_1) & (X_k - \mu_k)(X_2 - \mu_2) & \ldots & (X_k - \mu_k)(X_k - \mu_k) \end{bmatrix} \right\} \\ &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1k} \\ \vdots & \vdots & & \vdots \\ \sigma_{k1} & \sigma_{k2} & \ldots & \sigma_{kk} \end{bmatrix} = \Sigma\,. \end{aligned}$$

If $\mathbf{a} = (a_{1,}\ldots,a_k)'$, we see that:

$$\Sigma_{i=1}^{k} a_i X_i = \mathbf{a}'\mathbf{X}\,.$$

Basic properties of $\mathsf{E}(\mathbf{X})$ and $\mathsf{V}(\mathbf{X})$ are summarized by the following proposition.

4.1 Proposition *Let* $\mathbf{X} = (X_1,\ldots,X_k)'$ *a* $k \times 1$ *random vector,* α *a scalar,* **a** *and* **b** *fixed* $k \times 1$ *vectors, and* A *a fixed* $g \times k$ *matrix. Then, provided the moments considered are finite, we have the following properties:*

(a) $\mathsf{E}(\mathbf{X}+\mathbf{a}) = \mathsf{E}(\mathbf{X})+\mathbf{a}$;

(b) $\mathsf{E}(\alpha\mathbf{X}) = \alpha\mathsf{E}(\mathbf{X})$;

(c) $\mathsf{E}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\mathsf{E}(\mathbf{X})$, $\mathsf{E}(A\mathbf{X}) = A\mathsf{E}(\mathbf{X})$;

(d) $\mathsf{V}(\mathbf{X}+\mathbf{a}) = \mathsf{V}(\mathbf{X})$;

(e) $\mathsf{V}(\alpha\mathbf{X}) = \alpha^2\mathsf{V}(\mathbf{X})$;

(f) $\mathsf{V}(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\mathsf{V}(\mathbf{X})\,\mathbf{a}$, $\mathsf{V}(A\mathbf{X}) = A\mathsf{V}(\mathbf{X})A'$;

(g) $\mathsf{C}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \mathbf{a}'\mathsf{V}(\mathbf{X})\,\mathbf{b} = \mathbf{b}'\mathsf{V}(\mathbf{X})\,\mathbf{a}$.

4.2 Theorem *Let* $\mathbf{X} = (X_1, \ldots, X_k)'$ *be a random vector with covariance matrix* $\mathsf{V}(\mathbf{X}) = \Sigma$. *Then we have the following properties:*

(a) $\Sigma' = \Sigma$;

(b) Σ *is a positive semidefinite matrix;*

(c) $0 \leq |\Sigma| \leq \sigma_1^2\sigma_2^2\ldots\sigma_k^2$ *where* $\sigma_i^2 = \mathsf{V}(X_i)$, $i = 1, \ldots, k$;

(d) $|\Sigma| = 0 \Leftrightarrow$ *there is at least one linear relation between the* r.v.´s $X_1, \ldots, X_k$, *i.e., we can find constants* $a_1, \ldots, a_k$, b *not all equal to zero such that* $a_1X_1 + \cdots + a_kX_k = b$ *with probability 1;*

(e) *rank*$(\Sigma) = r < k \Leftrightarrow \mathbf{X}$ *can be expressed in the form*

$$\mathbf{X} = B\mathbf{Y} + \mathbf{c}$$

where $\mathbf{Y}$ *is a random vector of dimension* r *whose covariance matrix is* I_r, B *is a* $k \times r$ *matrix of rank* r*, and* $\mathbf{c}$ *is a* $k \times 1$ *constant vector.*

4.3 Remark We call the determinant $|\Sigma|$ the *generalized variance of* $\mathbf{X}$.

4.4 Definition *If we consider two random vectors* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *with dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively, the covariance matrix between* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *is defined by:*

$$\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) = \mathsf{E}\left\{[\mathbf{X}_1 - \mathsf{E}(\mathbf{X}_1)][\mathbf{X}_2 - \mathsf{E}(\mathbf{X}_2)]'\right\} .$$

The following proposition summarizes some basic properties of $\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2)$.

4.5 Proposition *Let* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *two random vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively. Then, provided the moments considered are finite we have the following properties:*

(a) $\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) = \mathsf{E}[\mathbf{X}_1 \mathbf{X}_2'] - \mathsf{E}(\mathbf{X}_1)\mathsf{E}(\mathbf{X}_2)'$;

(b) $\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) = \mathsf{C}(\mathbf{X}_2, \mathbf{X}_1)'$;

(c) $\mathsf{C}(\mathbf{X}_1, \mathbf{X}_1) = \mathsf{V}(\mathbf{X}_1)$, $\mathsf{C}(\mathbf{X}_2, \mathbf{X}_2) = \mathsf{V}(\mathbf{X}_2)$;

(d) *if* $\mathbf{a}$ *and* $\mathbf{b}$ *are fixed vectors of dimensions* $k_1 \times 1$ *and* $k_2 \times 1$ *respectively,*

$$\mathsf{C}(\mathbf{X}_1 + \mathbf{a}, \mathbf{X}_2 + \mathbf{b}) = \mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) ;$$

(e) *if* α *and* β *are two scalar constants,*

$$\mathsf{C}(\alpha \mathbf{X}_1, \beta \mathbf{X}_2) = \alpha\beta\, \mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) ;$$

(f) *if* $\mathbf{a}$ *and* $\mathbf{b}$ *are fixed* $k_1 \times 1$ *and* $k_2 \times 1$ *vectors,*

$$\mathsf{C}\left(\mathbf{a}'\mathbf{X}_1, \mathbf{b}'\mathbf{X}_2\right) = \mathbf{a}'\mathsf{C}\left(\mathbf{X}_1, \mathbf{X}_2\right)\mathbf{b} \ ;$$

(g) *if* A *and* B *are fixed matrices matrices with dimensions* $g_1 \times k_1$ *and* $g_2 \times k_2$ *respectively,*

$$\mathsf{C}\left(A\mathbf{X}_1, B\mathbf{X}_2\right) = \mathbf{A}\mathsf{C}\left(\mathbf{X}_1, \mathbf{X}_2\right)\mathbf{B}' \ ;$$

(h) *if* $k_1 = k_2$ *and* $\mathbf{X}_3$ *is a* $k \times 1$ *random vector,*

$$\mathsf{C}\left(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{X}_3\right) = \mathsf{C}\left(\mathbf{X}_1, \mathbf{X}_3\right) + \mathsf{C}\left(\mathbf{X}_2, \mathbf{X}_3\right) \ ;$$

(i) *if* $k_1 = k_2$,

$$\begin{aligned}
\mathsf{V}\left(\mathbf{X}_1 + \mathbf{X}_2\right) &= \mathsf{V}\left(\mathbf{X}_1\right) + \mathsf{V}\left(\mathbf{X}_2\right) + \mathsf{C}\left(\mathbf{X}_1, \mathbf{X}_2\right) + \mathsf{C}\left(\mathbf{X}_2, \mathbf{X}_1\right) \ , \\
\mathsf{V}\left(\mathbf{X}_1 - \mathbf{X}_2\right) &= \mathsf{V}\left(\mathbf{X}_1\right) + \mathsf{V}\left(\mathbf{X}_2\right) - \mathsf{C}\left(\mathbf{X}_1, \mathbf{X}_2\right) - \mathsf{C}\left(\mathbf{X}_2, \mathbf{X}_1\right) \ .
\end{aligned}$$