Multivariate distributions and measures of dependence between random variables *

Jean-Marie Dufour †
McGill University

September 2017
Compiled: September 5, 2017, 15:31

*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 Universitat 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
2. Covariances and correlations ............................... 2
3. Alternative interpretations of covariances and correlations .............................. 4
   3.1. Difference between two correlated random variables ......................... 5
   3.2. Polarization identity ........................................ 8
4. Covariance matrices ........................................... 8
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:

\[ C_t = \alpha + \beta Y_t + \epsilon_t \]

where \( \epsilon_t \) 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”:

\[ F_X (x) = \mathbb{P} [X \leq x] \]

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:

\[ \mu_X = \mathbb{E} (X) = \int_{-\infty}^{+\infty} x \, dF_X (x) \quad \text{(general case)} \]

\[ = \int_{-\infty}^{+\infty} x \, f_X (x) \, dx \quad \text{(continuous case)} \]

\[ \sigma_X^2 = \mathbb{V} (X) = \mathbb{E} [(X - \mu_X)^2] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \, dF_X (x) \quad \text{(general case)} \]

\[ = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \, f_X (x) \, dx \quad \text{(continuous case)} \]

\[ = \mathbb{E} (X^2) - [\mathbb{E} (X)]^2 \]

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:

\[ F(x, y) = \mathbb{P} [X \leq x, Y \leq y] \]
\[ \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) \, dx \, dy \]  

(continuous case.)

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:

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

\[
= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(x_1, x_2, \ldots, x_k) \, dx_1 \, dx_2 \cdots dx_k \]  

(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 \( \mu_X \) and \( \mu_Y \) and finite variances \( \sigma^2_X \) and \( \sigma^2_Y \). Below a.s. means “almost surely” (with probability 1).

2.1 Definition The covariance between \( X \) and \( Y \) is defined by

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

2.2 Definition Suppose \( \sigma^2_X > 0 \) and \( \sigma^2_Y > 0 \). Then the correlation between \( X \) and \( Y \) is defined by

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

When \( \sigma^2_X = 0 \) or \( \sigma^2_Y = 0 \), we set \( \rho_{XY} = 0 \).

2.3 Theorem The covariance and correlation between \( X \) and \( Y \) satisfy the following properties:

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

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

(c) \( \sigma_{XX} = \sigma^2_X \), \( \rho_{XX} = 1 \);

(d) \( \sigma^2_{XY} \leq \sigma^2_X \sigma^2_Y \) ; \hspace{1cm} (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 \( \sigma^2_X \neq 0 \) and \( \sigma^2_Y \neq 0 \),

\[ \rho^2_{XY} = 1 \Leftrightarrow \exists \text{ two constants } a \text{ and } b \text{ such that } a \neq 0 \text{ and } Y = aX + b \text{ a.s.} \]
\textbf{PROOF (a)}

\[
\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] \\
= E[XY - \mu_X Y - X \mu_Y + \mu_X \mu_Y] \\
= E(XY) - \mu_X E(Y) - E(X) \mu_Y + \mu_X \mu_Y \\
= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
= E(XY) - E(X)E(Y).
\]

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

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

for any arbitrary constant \( \lambda \). 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)

\[
\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2 \quad \Rightarrow \quad -\sigma_X \sigma_Y \leq \sigma_{XY} \leq \sigma_X \sigma_Y \\
\Rightarrow \quad -1 \leq \rho_{XY} \leq 1.
\]

(f)

\[
\sigma_{XY} = E\{(X - \mu_X)(Y - \mu_Y)\} = E(X - \mu_X)E(Y - \mu_Y) \\
= [E(X) - \mu_X][E(Y) - \mu_Y] = 0, \\
\rho_{XY} = \sigma_{XY} / \sigma_X \sigma_Y = 0.
\]
Note the reverse implication does not hold in general, i.e.,

\[ \rho_{XY} = 0 \not\implies X \text{ and } Y \text{ are independent} \]

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

\[ E(Y) = aE(X) + b = a\mu + b, \quad \sigma^2_Y = a^2 \sigma^2_X, \]

and

\[ \sigma_{XY} = E[(Y - \mu_Y)(X - \mu_X)] = E[a(X - \mu_X)(X - \mu_X)] = a\sigma^2_X. \]

Consequently,

\[ \rho_{XY}^2 = \frac{a^2 \sigma^4_X}{a^2 \sigma^2_X \sigma^2_X} = 1. \]

2) Sufficiency of the condition. If \( \rho_{XY}^3 = 1 \), then

\[ \sigma^2_{XY} - \sigma^2_X \sigma^2_Y = 0. \]

In this case, the equation

\[ E\left\{ \frac{1}{2} \left( (Y - \mu_Y) - \lambda (X - \mu_X) \right)^2 \right\} = \sigma^2_Y - 2\lambda \sigma_{XY} + \lambda^2 \sigma^2_X = 0 \]

has one and only one root

\[ \lambda = \frac{2\sigma_{XY}}{2\sigma^2_X} = \sigma_{XY}/\sigma^2_X, \]

so that

\[ E\left\{ \left( Y \sigma^2_Y - \mu_Y - \frac{\sigma_{XY}}{\sigma^2_X} (X - \mu_X) \right)^2 \right\} = 0 \]

and

\[ P\left[ Y = aX + b \right] = P\left[ Y = \frac{\sigma_{XY}}{\sigma^2_X} X + \left( \mu_Y - \frac{\sigma_{XY}}{\sigma^2_X} \mu_X \right) \right] = 1 \]

We can thus write:

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

where \( a = \sigma_{XY}/\sigma^2_X \) and \( b = \mu_Y - \frac{\sigma_{XY}}{\sigma^2_X} \mu_X \).

\[ \square \]

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

$$E[(Y - X)^2] = E\left\{\left(|(Y - \mu_Y) - (X - \mu_X)| - (\mu_Y - \mu_X)\right)^2\right\}$$

$$= E\left\{\left(|(Y - \mu_Y) - (X - \mu_X)|\right)^2\right\} + (\mu_Y - \mu_X)^2$$

$$= \sigma_Y^2 + \sigma_X^2 - 2\sigma_{XY} + (\mu_Y - \mu_X)^2$$

$$= \sigma_Y^2 + \sigma_X^2 - 2\rho_{XY}\sigma_X\sigma_Y + (\mu_Y - \mu_X)^2. \quad (3.1)$$

$E[(Y - X)^2]$ 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

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

and

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

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

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

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

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

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

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

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

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

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

$$E[(X + Y)^2] \leq \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \leq (\sigma_X + \sigma_Y)^2 + (\mu_X + \mu_Y)^2, \text{ if } \rho_{XY} \leq 0, \quad (3.10)$$
\[ E[(X+Y)^2] \geq \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 \geq (\sigma_X - \sigma_Y)^2 + (\mu_X + \mu_Y)^2, \quad \text{if } \rho_{XY} \geq 0, \tag{3.11} \]
\[ E[(Y+X)^2] = \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2, \quad \text{if } \rho_{XY} = 0. \tag{3.12} \]

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
\[ \tilde{X} = \frac{X - \mu_X}{\sigma_X}, \quad \tilde{Y} = \frac{Y - \mu_Y}{\sigma_Y}, \quad \rho(\tilde{X}, \tilde{Y}) = \rho(X,Y) = \rho_{XY}, \tag{3.14} \]
where we set \( \tilde{X} = 0 \) if \( \sigma_X = 0 \), and \( \tilde{Y} = 0 \) if \( \sigma_Y = 0 \). We then have:
\[ E(\tilde{X}) = E(\tilde{Y}) = 0, \quad V(\tilde{X}) = V(\tilde{Y}) = 1, \tag{3.15} \]
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
\[ E[(Y - X)^2] = E\left\{ (\mu_Y + \sigma_Y \tilde{Y} - (\mu_X + \sigma_X \tilde{X}))^2 \right\} \]
\[ = E\left\{ (\sigma_Y \tilde{Y} - \sigma_X \tilde{X})^2 + (\mu_Y - \mu_X)^2 \right\} \]
\[ = E\left\{ (\sigma_Y \tilde{Y} - \sigma_X \tilde{X})^2 \right\} + (\mu_Y - \mu_X)^2 \tag{3.18} \]

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

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

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

\[
\mu_Y = \mu_X \text{ and } \sigma_Y^2 = \sigma_X^2,
\] (3.23)
we have:

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

and

\[
0 \leq E[(Y - X)^2] \leq 4\sigma_X^2
\] (3.25)

so that

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

and, using Chebyshev’s inequality,

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

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

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

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

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

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

Since

\[
\sigma_Y(\bar{Y} - \bar{X}) = Y - \mu_Y - \frac{\sigma_Y}{\sigma_X}(X - \mu_X) = Y - \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) - \frac{\sigma_Y}{\sigma_X}X,
\] (3.32)

the linear function

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

can be viewed as a “forecast” of \(Y\) based on \(X\) such that

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

It is then of interest to note that

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

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

\[ V(X + Y) = V(X) + V(Y) + 2C(X, Y), \] (3.37)

\[ V(X - Y) = V(X) + V(Y) - 2C(X, Y), \] (3.38)

it is easy to see that

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

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

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

On \( X + Y \) and \( X - Y \), it also interesting to observe that

\[ C(X + Y, X - Y) = [V(X) - V(Y)] + [C(Y, X) - C(X, Y)] = V(X) - V(Y) \] (3.41)

so

\[ C((X + Y)/2, X - Y) = C(X + Y, X - Y) = 0, \quad \text{if } V(X) = V(Y). \] (3.42)

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

\[
\begin{align*}
E(X_i) &= \mu_i, \quad i = 1, \ldots, k, \\
C(X_i, X_j) &= \sigma_{ij}, \quad i, j = 1, \ldots, k.
\end{align*}
\]

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

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

\[ V \left[ \sum_{i=1}^{k} a_i X_i \right] = E \left\{ \left[ \sum_{i=1}^{k} a_i (X_i - \mu_i) \right] \left[ \sum_{j=1}^{k} a_j (X_j - \mu_j) \right] \right\} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \sigma_{ij}. \]

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 \( E(\mathbf{X}) \) by:

\[
\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}, \quad E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{pmatrix} \equiv \mu_X.
\]

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

\[
\mathbf{M} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}, \quad E(\mathbf{M}) = \begin{pmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{m1}) & E(X_{m2}) & \cdots & E(X_{mn}) \end{pmatrix}.
\]

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

\[
V(\mathbf{X}) = E\left\{ [\mathbf{X} - E(\mathbf{X})] [\mathbf{X} - E(\mathbf{X})]' \right\} = E\left\{ [\mathbf{X} - \mu_X] [\mathbf{X} - \mu_X]' \right\} = \begin{pmatrix} (X_1 - \mu_1)(X_1 - \mu_1) & \cdots & (X_1 - \mu_1)(X_k - \mu_k) \\ \vdots & \ddots & \vdots \\ (X_k - \mu_k)(X_1 - \mu_1) & \cdots & (X_k - \mu_k)(X_k - \mu_k) \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix} = \Sigma.
\]

If \( \mathbf{a} = (a_1, \ldots, a_k)' \), we see that:

\[
\sum_{i=1}^{k} a_i X_i = \mathbf{a}' \mathbf{X}.
\]

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

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

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

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

9
(c) \( E(a'X) = a'E(X), \; E(AX) = AE(X) \);

(d) \( V(X + a) = V(X) \);

(e) \( V(\alpha X) = \alpha^2 V(X) \);

(f) \( V(a'X) = a'V(X)a, \; V(AX) = AV(X)A' \);

(g) \( C(a'X, b'X) = a'V(X)b = b'V(X)a \).

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

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

(b) \( \Sigma \) is a positive semidefinite matrix;

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

(d) \( |\Sigma| = 0 \iff \) 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) \( \text{rank}(\Sigma) = r < k \iff X \) can be expressed in the form

\[
X = BY + c
\]

where \( Y \) is a random vector of dimension \( r \) whose covariance matrix is \( I_r \), \( B \) is a \( k \times r \) matrix of rank \( r \), and \( c \) is a \( k \times 1 \) constant vector.

4.3 Remark We call the determinant \( |\Sigma| \) the generalized variance of \( X \).

4.4 Definition If we consider two random vectors \( X_1 \) and \( X_2 \) with dimensions \( k_1 \times 1 \) and \( k_2 \times 1 \) respectively, the covariance matrix between \( X_1 \) and \( X_2 \) is defined by:

\[
C(X_1, X_2) = E \left\{ (X_1 - E(X_1)) (X_2 - E(X_2))' \right\}.
\]

The following proposition summarizes some basic properties of \( C(X_1, X_2) \).

4.5 Proposition Let \( X_1 \) and \( X_2 \) 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) \( C(X_1, X_2) = E[X_1X_2'] - E(X_1)E(X_2)' \);

(b) \( C(X_1, X_2) = C(X_2, X_1)' \);

(c) \( C(X_1, X_1) = V(X_1), \; C(X_2, X_2) = V(X_2) \).
(d) if \( \mathbf{a} \) and \( \mathbf{b} \) are fixed vectors of dimensions \( k_1 \times 1 \) and \( k_2 \times 1 \) respectively,
\[
C(X_1 + \mathbf{a}, X_2 + \mathbf{b}) = C(X_1, X_2) ;
\]

(e) if \( \alpha \) and \( \beta \) are two scalar constants,
\[
C(\alpha X_1, \beta X_2) = \alpha \beta C(X_1, X_2) ;
\]

(f) if \( \mathbf{a} \) and \( \mathbf{b} \) are fixed \( k_1 \times 1 \) and \( k_2 \times 1 \) vectors,
\[
C(a' X_1, b' X_2) = a' C(X_1, X_2) b ;
\]

(g) if \( A \) and \( B \) are fixed matrices matrices with dimensions \( g_1 \times k_1 \) and \( g_2 \times k_2 \) respectively,
\[
C(A X_1, B X_2) = A C(X_1, X_2) B' ;
\]

(h) if \( k_1 = k_2 \) and \( \mathbf{X}_3 \) is a \( k \times 1 \) random vector,
\[
C(X_1 + X_2, \mathbf{X}_3) = C(X_1, \mathbf{X}_3) + C(X_2, \mathbf{X}_3) ;
\]

(i) if \( k_1 = k_2 \),
\[
\begin{align*}
\nabla (X_1 + X_2) &= \nabla (X_1) + \nabla (X_2) + C(X_1, X_2) + C(X_2, X_1), \\
\nabla (X_1 - X_2) &= \nabla (X_1) + \nabla (X_2) - C(X_1, X_2) - C(X_2, X_1).
\end{align*}
\]