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
This version: September 2011
Compiled: September 29, 2011, 13:26

* This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), 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 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; 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":*

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

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:

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

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

$$\mathsf{V}(X) = \sigma_X^2 = \mathsf{E}\left[(X-\mu_X)^2\right] = \int_{-\infty}^{+\infty} (x-\mu_X)^2\, dF_X(x) \qquad \text{(general case)}$$

$$= \int_{-\infty}^{+\infty} (x-\mu_X)^2\, F_X(x)\, dx \qquad \text{(continuous case)}$$

$$= \mathsf{E}\left(X^2\right) - [\mathsf{E}(X)]^2$$

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:

$$F(x, y) \quad = \quad \mathsf{P}[X \leq x,\, Y \leq y]$$

$$= \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy \qquad \text{(continuous case.)}$$

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

$$F(x_1, x_2, \ldots, x_k) = \mathsf{P}[X_1 \le x_1, X_2 \le x_2, \ldots, X_k \le x_k]$$
$$= \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f(x_1, x_2, \ldots, x_k)\, dx_1 dx_2 \cdots dx_k \qquad \text{(continuous case)}$$

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

2.1. Covariance and correlation between two random variables

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*

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

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

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

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

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

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

PROOF (a)

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

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

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

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)

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

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

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

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 $b = \mu_Y - \frac{\sigma_{XY}}{\sigma_y^2}\mu_X$. □

2.2. Covariances and correlations between k random variables

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

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

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.

2.4 Proposition *Let* $\mathbf{X} = (X_1, \ldots, X_k)'$ *a* $k \times 1$ *random vector,* α *a scalar,* $\mathbf{a}$ *and* $\mathbf{b}$ *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) $\mathsf{C}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \mathbf{a}'\mathsf{V}(\mathbf{X})\mathbf{b} = \mathbf{b}'\mathsf{V}(\mathbf{X})\mathbf{a}$.

2.5 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) $0 \le |\Sigma| \le \sigma_1^2\sigma_2^2 \ldots \sigma_k^2$ *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.*

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

2.7 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 $\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2)$.

2.8 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,*

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

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

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

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

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

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

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

3. Multinormal distribution

Consider two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ with dimensions $k_1 \times 1$ and $k_2 \times 1$ respectively. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then

$$\mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) \equiv \mathsf{E}\left[\left(\mathbf{X}_1 - \mu_{X_1}\right)\left(\mathbf{X}_2 - \mu_{X_2}\right)'\right] = 0$$

The reverse implication is not true in general, except in special cases. One such case is the one where the random vector $\mathbf{X} = \left(\mathbf{X}_1', \mathbf{X}_2'\right)'$ follows a multinormal distribution.

3.1 Definition *We say that the* $k \times 1$ *random vector* $\mathbf{X}$ *follows a multinormal distribution with mean*

μ and covariance matrix Σ, denoted $\mathbf{X} \sim N_k[\mu, \Sigma]$, if the characteristic function of $\mathbf{X}$ has the form:

$$\mathsf{E}\left[e^{i\mathbf{t}'\mathbf{X}}\right] = e^{i\mu'\mathbf{t} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}, \quad \mathbf{t} \in \mathscr{R}^k, \; i = \sqrt{-1}\,.$$

3.2 When $|\Sigma| \neq 0$, the vector $\mathbf{X}$ has a density function of the form:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right]$$

If $k = 1$, then $\Sigma = \sigma^2$ and

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}(x-\mu)\frac{1}{\sigma^2}(x-\mu)\right] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right].$$

Some important properties of the multinormal distribution are summarized in the following theorem.

3.3 Theorem *If $\mathbf{X} \sim N_k[\mu, \Sigma]$, then*

(a) $\mathbf{X} + \mathbf{c} \sim N_k[\mu + \mathbf{c}, \Sigma]$ *, for any fixed $k \times 1$ vector $\mathbf{c}$;*

(b) $\mathbf{a}'\mathbf{X} \sim N_1[\mathbf{a}'\mu, \mathbf{a}'\Sigma\mathbf{a}]$ *, for any fixed $k \times 1$ vector $\mathbf{a}$;*

(c) $A\mathbf{X} \sim N_g[A\mu, A\Sigma A']$ *, for any fixed $g \times k$ matrix A ;*

(d) *if*

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_k\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right],$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are vectors of dimensions $k_1 \times 1$ and $k_2 \times 1$,

$$\begin{aligned} \mu_1 &= \mathsf{E}(\mathbf{X}_1), \mu_2 = \mathsf{E}(\mathbf{X}_2), \; \Sigma_{11} = \mathsf{C}(\mathbf{X}_1, \mathbf{X}_1), \; \Sigma_{22} = \mathsf{C}(\mathbf{X}_2, \mathbf{X}_2), \\ \Sigma_{12} &= \mathsf{C}(\mathbf{X}_1, \mathbf{X}_2) = \Sigma_{21}', \end{aligned}$$

then

(i) $\mathbf{X}_1 \sim N_{k_1}[\mu_1, \Sigma_{11}]$, $\mathbf{X}_2 \sim N_{k_2}[\mu_2, \Sigma_{22}]$;

(ii) *$\mathbf{X}_1$ and $\mathbf{X}_2$ are independent $\Leftrightarrow \Sigma_{12} = 0$;*

(iii) *the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1$ is normal with mean and et variance*

$$\begin{aligned} \mathsf{E}[\mathbf{X}_2 | \mathbf{X}_1] &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \mu_1), \\ \mathsf{V}[\mathbf{X}_2 | \mathbf{X}_1] &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}, \end{aligned}$$

i.e.

$$\mathbf{X}_2 | \mathbf{X}_1 \sim N_{k_2}\left[\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right].$$

3.4 Theorem *If* $\mathbf{X} \sim N_k[\mu, \Sigma]$ *with* $|\Sigma| \neq 0$, *then*

$$(\mathbf{X}-\mu)' \Sigma^{-1} (\mathbf{X}-\mu) \sim \chi^2(k) \ .$$

PROOF Since Σ is a positive definite matrix ($|\Sigma| \neq 0$), there exists a nonsingular matrix P such that

$$P\Sigma P' = I_k$$

hence

$$\begin{aligned} \Sigma &= P^{-1}(P')^{-1} = (P'P)^{-1} \ , \\ \Sigma^{-1} &= P'P \ . \end{aligned}$$

Consequently,

$$\begin{aligned} (\mathbf{X}-\mu)' \Sigma^{-1} (\mathbf{X}-\mu) &= (\mathbf{X}-\mu)' P'P (\mathbf{X}-\mu) \\ &= [P(\mathbf{X}-\mu)]' [P(\mathbf{X}-\mu)] = \mathbf{v}'\mathbf{v} = \Sigma_{i=1}^k v_i^2 \end{aligned}$$

where

$$\mathbf{v} \equiv P[\mathbf{X}-\mu] = (v_1, v_2, \ldots, v_k)' \ .$$

Since $\mathbf{X} \sim N[\mu, \Sigma]$, we have $\mathbf{X}-\mu \sim N[\mathbf{0}, \Sigma]$, hence

$$P[\mathbf{X}-\mu] \sim N\left[\mathbf{0}, P\Sigma P'\right] \ ,$$

and

$$\mathbf{v} = P[\mathbf{X}-\mu] \sim N[\mathbf{0}, I_k] \ .$$

Thus $v_1, \ldots, v_k$ are i.i.d. $N[0,1]$ and $(\mathbf{X}-\mu)' \Sigma^{-1} (\mathbf{X}-\mu) = \Sigma_{i=1}^k v_i^2 \sim \chi^2(k)$. □