Covariance, correlation and linear regression between random variables *

Jean-Marie Dufour †
McGill University

January 15, 2025

–

* 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), 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

List of Definitions, Assumptions, Propositions and Theorems iv

1. Random variables 1

2. Covariances and correlations 2

3. Regression coefficients between two variables 8

4. Uncentered covariances, correlations and regression coefficients 9

5. Difference and sum of two correlated random variables 9
   - 5.1. Uncentered second moments . . . 10
   - 5.2. Covariances . . . 10
   - 5.3. Correlations . . . 12
   - 5.4. Inequalities . . . 13
   - 5.5. Polarization identities . . . 16

6. Hoeffding representation 17

7. Linear regression and correlations 23
   - 7.1. Linear approximation . . . 23
   - 7.2. Regression coefficients as solutions of moment equations . . . 25
   - 7.3. Decompositions . . . 25
   - 7.4. Population t and F coefficients . . . 27
   - 7.5. Inequalities on linear regression coefficients . . . 29

8. Covariance and variance decompositions **30**
8.1. Sum of random variables . . . 30
8.1.1. Covariance decomposition . . . 30
8.1.2. Covariance-variance decomposition . . . 36
8.1.3. Covariance-variance subdecompositions . . . 39
8.2. Linear combination of random variables . . . 41
8.3. Linear combination of random variables with disturbance . . . 44
8.4. Factor decompositions . . . 45

9. Sources and additional references **47**

List of Definitions, Assumptions, Propositions and Theorems

Definition 2.1 : Covariance . . . 3
Definition 2.2 : Correlation . . . 3
Theorem 2.1 : Basic properties of covariances and correlations . . . 3
Theorem 2.2 : Identification of linear transformation of a random variable . . . 6
Definition 3.1 : Linear regression coefficient . . . 8
Definition 4.1 : Uncentered covariance . . . 9
Definition 4.2 : Uncentered correlation . . . 9
Definition 4.3 : Uncentered linear regression coefficient . . . 9
Lemma 6.1 : Indicator representation of continuous variable . . . 17
Proof of Lemma 6.1 . . . 18
Theorem 6.2 : Hoeffding covariance identity . . . 19
Proof of Theorem 6.2 . . . 20
Corollary 6.3 : Hoeffding-type variance representation . . . 21
Corollary 6.4 : Hoeffding-type conditional distribution representation . . . 21
Corollary 6.5 : Hoeffding covariance identit:y location invariance . . . 21
Proposition 6.6 : Copula representation of covariance . . . 22
Proposition 7.1 : Mean optimality . . . 23
Proposition 7.2 : Uncentered regression optimality . . . 23
Proposition 7.3 : Centered regression optimality . . . 24

1. Random variables

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".

Definition 1.1 *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":*

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

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:

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

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

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

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

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

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:

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

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:

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

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

$$\mu_X := \mathbb{E}(X), \quad \mu_Y := \mathbb{E}(Y), \tag{2.1}$$

and finite second moments

$$\bar{\sigma}_X^2 := \mathbb{E}(X^2), \quad \bar{\sigma}_Y^2 := \mathbb{E}(Y^2). \tag{2.2}$$

Then, X and Y have finite variances:

$$\sigma_X^2 := \mathbb{V}(X) := \mathbb{E}[(X-\mu_X)^2] = \mathbb{E}(X^2) - \mu_X^2 = \bar{\sigma}_X^2 - \mu_X^2, \tag{2.3}$$

$$\sigma_Y^2 := \mathbb{V}(Y) := \mathbb{E}[(Y-\mu_Y)^2] = \mathbb{E}(Y^2) - \mu_Y^2 = \bar{\sigma}_Y^2 - \mu_Y^2. \tag{2.4}$$

We also denote:

$$\bar{\sigma}(X) := \bar{\sigma}_X = [\mathbb{E}(X^2)]^{1/2}, \quad \bar{\sigma}(Y) := \bar{\sigma}_Y = [\mathbb{E}(Y^2)]^{1/2}, \tag{2.5}$$

$$\sigma(X) := \sigma_X, \quad \sigma(Y) := \sigma_Y, \tag{2.6}$$

where $\bar{\sigma}(X) \geq 0$, $\bar{\sigma}(Y) \geq 0$, $\sigma(X) \geq 0$ and $\sigma(Y) \geq 0$, so that

$$\sigma(X)^2 = \mathbb{V}(X), \quad \sigma(Y)^2 = \mathbb{V}(Y). \tag{2.7}$$

Below *a.s.* means "almost surely" (with probability 1). In particular, we have:

$$\mathbb{E}(X^2) = 0 \Leftrightarrow [X = 0 \quad \text{a.s.}] \Leftrightarrow \mathbb{P}[X = 0] = 1\,, \tag{2.8}$$

$$\mathbb{V}(X) = 0 \Leftrightarrow [X = \mathbb{E}(X) \quad \text{a.s.}] \Leftrightarrow \mathbb{P}[X = \mathbb{E}(X)] = 1\,. \tag{2.9}$$

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

$$\mathsf{C}(X, Y) := \sigma_{XY} := \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right]\,. \tag{2.10}$$

When $\mathsf{C}(X, Y) = 0$, *we say that X and Y are* orthogonal.

Definition 2.2 CORRELATION. *The* correlation *between X and Y is defined by*

$$\rho(X, Y) := \rho_{XY} := \frac{\mathsf{C}(X, Y)}{\sigma(X)\sigma(Y)} \tag{2.11}$$

where we set $\rho(X, Y) := 0$ *when* $\sigma(X)\sigma(Y) = 0$.

When X or Y is degenerate, we have $\mathsf{C}(X, Y) = \sigma(X)\sigma(Y) = 0$. The convention $\rho(X, Y) := 0$ when $\sigma(X)\sigma(Y) = 0$ is motivated by the fact that $\mathsf{C}(X, Y) = 0$ in this case.

Theorem 2.1 BASIC PROPERTIES OF COVARIANCES AND CORRELATIONS. *Let* (X, Y) *be a pair of random variables with finite second moments. The covariance and correlation between X and Y satisfy the following properties:*

(*a*) $\mathsf{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;

(*b*) $\mathsf{C}(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 \mathsf{C}(X, Y)$ *for any constants* a_1, a_2, b_1, b_2 ;

(*c*) $\rho(a_1 + b_1 X, a_2 + b_2 Y) = \rho(X, Y)$ *for any constants* a_1, a_2, b_1, b_2 *such that* $b_1 b_2 \neq 0$;

(*d*) $\mathsf{C}(X, Y) = \mathsf{C}(Y, X)$ *and* $\rho(X, Y) = \rho(Y, X)$;

(*e*) $\mathsf{C}(X, X) = \mathbb{V}(X)$;

(*f*) $\rho(X, X) = 1$ *if* $\mathbb{V}(X) > 0$;

(*g*) $\mathsf{C}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y)$; (*Cauchy-Schwarz inequality*)

(*h*) $-1 \leq \rho(X, Y) \leq 1$;

(i) *X and Y are independent* $\Rightarrow \mathsf{C}(X,Y) = 0 \Rightarrow \rho(X,Y) = 0$;

(j) *if* $\sigma(X)\sigma(Y) \neq 0$, *then*

$$\begin{aligned}[\rho(X,Y)^2 = 1] &\Leftrightarrow \left[\exists \text{ two constants } a \text{ and } b \text{ such that } b \neq 0 \text{ and } Y = a + bX \text{ a.s.}\right] \\ &\Leftrightarrow \left[Y = a + bX \text{ a.s. with } b = \beta(X \rightharpoondown Y) \text{ and } a = \mathbb{E}(Y) - b\,\mathbb{E}(X)\right], \end{aligned} \tag{2.12}$$

$$[\rho(X,Y) = 1] \Leftrightarrow \left[Y = a + bX \text{ a.s. with } b = \beta(X \rightharpoondown Y) > 0 \text{ and } a = \mathbb{E}(Y) - b\,\mathbb{E}(X)\right], \tag{2.13}$$

$$[\rho(X,Y) = -1] \Leftrightarrow \left[Y = a + bX \text{ a.s. with } b = \beta(X \rightharpoondown Y) < 0 \text{ and } a = \mathbb{E}(Y) - b\,\mathbb{E}(X)\right]. \tag{2.14}$$

PROOF (a)

$$\begin{aligned}\mathsf{C}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathbb{E}[XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y] \\ &= \mathbb{E}(XY) - \mu_X\mathbb{E}(Y) - \mathbb{E}(X)\mu_Y + \mu_X\mu_Y \\ &= \mathbb{E}(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\ . \end{aligned} \tag{2.15}$$

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

$$\begin{aligned}\mathbb{E}\left\{[Y-\mu_Y-\lambda(X-\mu_X)]^2\right\} &= \mathbb{E}\left\{[(Y-\mu_Y)-\lambda(X-\mu_X)]^2\right\} \\ &= \mathbb{E}\left\{(Y-\mu_Y)^2 - 2\lambda(X-\mu_X)(Y-\mu_Y) + \lambda^2(X-\mu_X)^2\right\} \\ &= \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 \geq 0 \end{aligned} \tag{2.16}$$

for any arbitrary constant λ. In other words, the second-order polynomial

$$g(\lambda) = \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 \tag{2.17}$$

cannot take negative values. This can happen only if the equation

$$\lambda^2\sigma_X^2 - 2\lambda\sigma_{XY} + \sigma_Y^2 = 0 \tag{2.18}$$

does not have two distinct real roots, *i.e.* the roots are either complex or identical. The roots of equation (2.18) are:

$$\lambda = \frac{2\sigma_{XY} \pm \sqrt{4\sigma_{XY}^2 - 4\sigma_X^2\sigma_Y^2}}{2\sigma_X^2} = \frac{\sigma_{XY} \pm \sqrt{\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2}}{\sigma_X^2} \,. \tag{2.19}$$

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

$$\sigma_{XY}^2 \le \sigma_X^2\sigma_Y^2 \,. \tag{2.20}$$

(h)

$$\begin{aligned} \sigma_{XY}^2 \le \sigma_X^2\sigma_Y^2 \quad &\Rightarrow \quad -\sigma_X\sigma_Y \le \sigma_{XY} \le \sigma_X\sigma_Y \\ &\Rightarrow \quad -1 \le \rho_{XY} \le 1 \,. \end{aligned} \tag{2.21}$$

(i) If X and Y are independent, we have:

$$\begin{aligned} \sigma_{XY} &= \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\} = \mathbb{E}(X-\mu_X)\,\mathbb{E}(Y-\mu_Y) \\ &= [\mathbb{E}(X)-\mu_X][\mathbb{E}(Y)-\mu_Y] = 0\,, \end{aligned} \tag{2.22}$$

$$\rho_{XY} = \sigma_{XY}\big/\sigma_X\sigma_Y = 0\,. \tag{2.23}$$

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

$$\rho_{XY} = 0 \not=> X \text{ and } Y \text{ are independent}\,. \tag{2.24}$$

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

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b = a\mu_X + b\,,\ \ \sigma_Y^2 = a^2\sigma_X^2\,, \tag{2.25}$$

and

$$\sigma_{XY} = \mathbb{E}[(Y-\mu_Y)(X-\mu_X)] = \mathbb{E}[a(X-\mu_X)(X-\mu_X)] = a\sigma_X^2\,. \tag{2.26}$$

Consequently,

$$\rho_{XY}^2 = \frac{a^2\sigma_X^4}{a^2\sigma_X^2\sigma_X^2} = 1\,. \tag{2.27}$$

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

$$\sigma_{XY}^2 - \sigma_X^2\sigma_Y^2 = 0. \tag{2.28}$$

In this case, the equation

$$\mathbb{E}\left\{[(Y-\mu_Y) - \lambda(X-\mu_X)]^2\right\} = \sigma_Y^2 - 2\lambda\sigma_{XY} + \lambda^2\sigma_X^2 = 0 \tag{2.29}$$

has one and only one root

$$\lambda = \frac{2\sigma_{XY}}{2\sigma_X^2} = \sigma_{XY}/\sigma_X^2 \,, \tag{2.30}$$

so that

$$\mathbb{E}\{[\left(Y\sigma_Y^2 - \mu_Y\right) - (\sigma_{XY}/\sigma_X^2)\,(X-\mu_X)]^2\} = 0 \tag{2.31}$$

and

$$\mathbb{P}\left[(Y-\mu_Y) - (\sigma_{XY}/\sigma_X^2)\,(X-\mu_X) = 0\right] = \mathbb{P}\left[Y = \left(\mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X\right) + (\sigma_{XY}/\sigma_X^2)X\right] = 1 \tag{2.32}$$

We can thus write:

$$Y = a + bX \text{ with probability 1} \tag{2.33}$$

where $b = \sigma_{XY}/\sigma_X^2$ and $a = \mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X$. This establishes (2.12). (2.13) follows on observing that, for $b = \sigma_{XY}/\sigma_X^2$ and $a = \mu_Y - (\sigma_{XY}/\sigma_X^2)\mu_X$,

$$\begin{aligned}
[\rho\,(X,Y) = 1] \quad &\Leftrightarrow \quad \left\{\rho\,(X,Y)^2 = 1 \text{ and } \rho\,(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a+bX]\right] = 1 \text{ and } \rho\,(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a+bX] = 1 \text{ and } \rho\,(X,Y) = \frac{b\sigma_X^2}{\sigma_X\sigma_Y} > 0\right\} \\
&\Leftrightarrow \quad \left\{\mathbb{P}[Y = a+bX] = 1 \text{ and } b > 0\right\}.
\end{aligned} \tag{2.34}$$

The proof for (2.14) is similar. □

A basic problem in this context consists in considering the case where

$$Y = a + bX \quad a.s. \tag{2.35}$$

and find whether a and b can be determined (or "identified") from the first and second moments of X and Y. The following theorem shows that a and b are uniquely determined if only if $\mathbb{V}(X) > 0$.

Theorem 2.2 IDENTIFICATION OF LINEAR TRANSFORMATION OF A RANDOM VARIABLE. *Suppose X and Y satisfy the linear equation* (2.35). *If* $\mathbb{V}(X) > 0$, *then*

$$\{\mathbb{P}[Y = a_1 + b_1 X] = 1\} \Rightarrow [a_1 = a \text{ and } b_1 = b]\,. \tag{2.36}$$

If $\mathbb{V}(X) = 0$, *then, for all* $b_1 \in \mathbb{R}$,

$$\mathbb{P}[Y = a^* + b_1 X] = 1 \tag{2.37}$$

where $a^* = \mathbb{E}(Y) - b_1\mathbb{E}(X)$.

PROOF By (2.35), we have

$$\mathbb{E}(Y) = a + b\mathbb{E}(X)\,. \tag{2.38}$$

Suppose $\mathbb{P}[Y = a_1 + b_1 X] = 1$ holds. Then

$$Y = a_1 + b_1 X = a + bX \quad \text{a.s.} \tag{2.39}$$

hence

$$(a_1 - a) + (b_1 - b)X = 0 \quad \text{a.s.} \tag{2.40}$$

$$\mathbb{V}[(a_1 - a) + (b_1 - b)X] = \mathbb{V}[(b_1 - b)X] = (b_1 - b)^2\mathbb{V}(X) = 0\,. \tag{2.41}$$

If $\mathbb{V}(X) > 0$, this entails $b_1 = b$, which in turn implies

$$Y = a_1 + b_1 X = a_1 + bX = a + bX \tag{2.42}$$

hence $a_1 = a$. If $\mathbb{V}(X) = 0$, then

$$X - \mathbb{E}(X) = 0 \quad \text{a.s.}, \tag{2.43}$$

hence, for any $b_1 \in \mathbb{R}$,

$$b[X - \mathbb{E}(X)] = b_1[X - \mathbb{E}(X)] = 0 \text{ a.s.}, \tag{2.44}$$

and

$$\begin{aligned} Y &= a + bX \\ &= a + b\mathbb{E}(X) + b[X - \mathbb{E}(X)] \\ &= \mathbb{E}(Y) + b_1[X - \mathbb{E}(X)] \\ &= [\mathbb{E}(Y) - b_1\mathbb{E}(X)] + b_1 X \\ &= a^* + b_1 X \quad \text{a.s.} \end{aligned} \tag{2.45}$$

where $a^* := [\mathbb{E}(Y) - b_1\mathbb{E}(X)]$. □

If $\mathbb{V}(X) > 0$, there is only one pair (a, b) which satisfies (2.35). If $\mathbb{V}(X) = 0$, Y has several representations of the form $a + bX$: the values a and b are not "identified". But they are not completely undetermined. Once b is specified, a is determined by the equation

$$a = \mathbb{E}(Y) - b\mathbb{E}(X)\,. \tag{2.46}$$

Indeed, if (2.39) holds, we must have

$$(b_1 - b)\mathbb{E}(X) = a - a_1 \,. \tag{2.47}$$

Corollary 2.3 *Under the assumptions of Theorem* 2.1*,*

$$[\rho(X,Y)^2 = 1] \Leftrightarrow \left[\exists \text{ two unique constants } a \text{ and } b \text{ such that } b \neq 0 \text{ and } Y = a + bX \text{ a.s.}\right].$$

3. Regression coefficients between two variables

Definition 3.1 LINEAR REGRESSION COEFFICIENT. *The* linear regression coefficient *of Y on X is defined by*

$$\beta(X \rightharpoondown Y) := \frac{\mathsf{C}(X,Y)}{\mathbb{V}(X)} \tag{3.1}$$

where we set $\beta(X \rightharpoondown Y) := 0$ when $\mathbb{V}(X) = 0$. By convention,

$$\beta(Y \leftharpoondown X) = \beta(X \rightharpoondown Y) \,. \tag{3.2}$$

The "harpoon" symbols $\rightharpoondown$ and $\leftharpoondown$ represent a statistical "dependence" or "predictability" relation; for example, $X \rightharpoondown Y$ and $Y \leftharpoondown X$ represent dependence of Y on X. The relation $X \rightharpoondown Y$ is typically asymmetric: $Y \leftharpoondown X$ represents a different relation. It does not necessarily correspond to a "causal" relation. From the above definitions, we can write:

$$\mathsf{C}(X,Y) = \rho(X,Y)\,\sigma(X)\,\sigma(Y) \tag{3.3}$$

which holds in all cases [including when $\sigma(X) = 0$ or $\sigma(Y) = 0$]. When $\sigma(X) > 0$, we also have:

$$\beta(X \rightharpoondown Y) = \frac{\rho(X,Y)\,\sigma(X)\sigma(Y)}{\sigma(X)^2} = \rho(X,Y)\,\frac{\sigma(Y)}{\sigma(X)} \,. \tag{3.4}$$

When $\sigma(X) > 0$, we have [by (3.4) and Theorem 2.1(h)]:

$$-\frac{\sigma(Y)}{\sigma(X)} \leq \beta(X \rightharpoondown Y) = \rho(X,Y)\,\frac{\sigma(Y)}{\sigma(X)} \leq \frac{\sigma(Y)}{\sigma(X)} \tag{3.5}$$

so that the regression coefficient can be bounded the variance ratio $\sigma(Y)/\sigma(X)$. More generally, if $\sigma(X) > 0$ and

$$\rho_L \leq \rho(X,Y) \leq \rho_U \,, \tag{3.6}$$

we have

$$\rho_L \frac{\sigma(Y)}{\sigma(X)} \leq \beta(X \rightharpoondown Y) \leq \rho_U \frac{\sigma(Y)}{\sigma(X)} . \tag{3.7}$$

4. Uncentered covariances, correlations and regression coefficients

Definition 4.1 UNCENTERED COVARIANCE. *The* uncentered covariance *between X and Y is defined by*

$$\bar{\mathsf{C}}(X,Y) := \bar{\sigma}_{XY} := \mathbb{E}[XY] . \tag{4.1}$$

When $\bar{\mathsf{C}}(X,Y) = 0$, we say that X and Y are orthogonal with respect to zero.

Definition 4.2 UNCENTERED CORRELATION. *The* uncentered correlation *between X and Y is defined by*

$$\bar{\rho}(X,Y) := \bar{\rho}_{XY} := \frac{\bar{\mathsf{C}}(X,Y)}{\bar{\sigma}(X)\bar{\sigma}(Y)} \tag{4.2}$$

where we set $\rho(X,Y) := 0$ when $\bar{\sigma}(X)\bar{\sigma}(Y) = 0$.

Definition 4.3 UNCENTERED LINEAR REGRESSION COEFFICIENT. *The* uncentered linear regression coefficient *of Y on X is defined by*

$$\bar{\beta}(X \rightharpoondown Y) := \frac{\bar{\mathsf{C}}(X,Y)}{\bar{\sigma}(X)} \tag{4.3}$$

where we set $\bar{\beta}(X \rightharpoondown Y) := 0$ when $\bar{\sigma}(X) = 0$.

5. Difference and sum of two correlated random variables

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. through a “decoupling” representation of covariances and correlations;

4. Hoeffding identity;

5. by looking at the linear regression of Y on X;

5.1. Uncentered second moments

Let us look the difference and the sum of two random variables X and Y:

$$\mathbb{E}[(Y-X)^2] = \mathbb{E}(X^2+Y^2-2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)-2\mathbb{E}(XY)\,. \tag{5.1}$$

$$\mathbb{E}[(Y+X)^2] = \mathbb{E}(X^2+Y^2+2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)+2\mathbb{E}(XY)\,. \tag{5.2}$$

From these, we see that:

$$\mathbb{E}(XY) = \frac{1}{2}\{[\mathbb{E}(X^2)+\mathbb{E}(Y^2)]-\mathbb{E}[(Y-X)^2]\}\,, \tag{5.3}$$

$$\mathbb{E}(XY) = \frac{1}{2}\{\mathbb{E}[(Y+X)^2]-[\mathbb{E}(X^2)+\mathbb{E}(Y^2)]\,. \tag{5.4}$$

The cross second moment $\mathbb{E}(XY)$ can be interpreted in two ways in terms of (uncentered) second moments:

1. $\mathbb{E}(XY)$ is equal to half the difference between the sum of the second moments X and Y and the second moment of $Y-X$;

2. $\mathbb{E}(XY)$ is equal to half the difference between the second moment of $Y+X$ and the sum of the second moments of X and Y.

5.2. Covariances

We now consider similar expressions for the covariance $\sigma_{XY} = \mathbb{E}[(Y-\mu_Y)-(X-\mu_X)]$. It is easy to see that

$$\begin{aligned}\mathbb{E}[(Y-X)^2] &= \mathbb{E}\left\{\left([(Y-\mu_Y)-(X-\mu_X)]+(\mu_Y-\mu_X)\right)^2\right\} \\ &= \mathbb{E}\{[(Y-\mu_Y)-(X-\mu_X)]^2\}+(\mu_Y-\mu_X)^2 \\ &= \sigma_Y^2+\sigma_X^2-2\sigma_{XY}+(\mu_Y-\mu_X)^2 \\ &= \sigma_Y^2+\sigma_X^2-2\rho_{XY}\sigma_X\sigma_Y+(\mu_Y-\mu_X)^2\,.\end{aligned} \tag{5.5}$$

$\mathbb{E}[(Y-X)^2]$ 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 (5.5) shows clearly that $\mathbb{E}[(Y-X)^2]$ tends to be large, when Y and X very different means or variances. Similarly,

$$\begin{aligned}
\mathbb{E}[(Y+X)^2] &= \mathbb{E}\left\{\left([(Y-\mu_Y)+(X-\mu_X)]+(\mu_Y+\mu_X)\right)^2\right\} \\
&= \mathbb{E}\left\{[(Y-\mu_Y)+(X-\mu_X)]^2\right\}+(\mu_Y+\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2+2\sigma_{XY}+(\mu_Y+\mu_X)^2 \\
&= \sigma_Y^2+\sigma_X^2+2\rho_{XY}\sigma_X\sigma_Y+(\mu_Y+\mu_X)^2 \,. \qquad (5.6)
\end{aligned}$$

From (5.5), we see that

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{(\sigma_Y^2+\sigma_X^2)-\mathbb{E}[(Y-X)^2]+(\mu_Y-\mu_X)^2\} \\
&= \frac{1}{2}\left[(\sigma_Y^2+\sigma_X^2)-\mathbb{E}\{[(Y-\mu_Y)-(X-\mu_X)]^2\}\right] \\
&= \frac{1}{2}\left[(\sigma_Y^2+\sigma_X^2)-\mathbb{V}(Y-X)\right] \\
&= \frac{1}{2}\left[\mathbb{V}(Y)+\mathbb{V}(X)-\mathbb{V}(Y-X)\right]. \qquad (5.7)
\end{aligned}$$

σ_{XY} represents the difference between the sum of the variances of X and Y and the variance of $Y-X$. In particular, if $\mu_Y = \mu_X$,

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{\sigma_Y^2+\sigma_X^2-\mathbb{E}[(Y-X)^2]\}. \\
&= \frac{1}{2}\{\mathbb{V}(Y)+\mathbb{V}(X)-\mathbb{E}[(Y-X)^2]\}\,. \qquad (5.8)
\end{aligned}$$

In this case, σ_{XY} represents the difference between the sum of the variances of X and Y and the mean square difference $\mathbb{E}[(Y-X)^2$.

Similarly, by (5.6), we have:

$$\begin{aligned}
\sigma_{XY} &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2]-(\sigma_Y^2+\sigma_X^2)-(\mu_Y+\mu_X)^2\} \\
&= \frac{1}{2}\left[\mathbb{E}\{[(Y-\mu_Y)+(X-\mu_X)]^2\}-(\sigma_Y^2+\sigma_X^2)\right] \\
&= \frac{1}{2}\left[\mathbb{V}(Y+X)-(\sigma_Y^2+\sigma_X^2)\right]
\end{aligned}$$

$$= \frac{1}{2}\left[\mathbb{V}(Y+X) - [\mathbb{V}(Y)+\mathbb{V}(X)]\right]. \tag{5.9}$$

σ_{XY} represents the difference between the variance of $Y+X$ and the sum of the variances of X and Y. In particular, if $\mu_Y = \mu_X = 0$,

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - (\sigma_Y^2 + \sigma_X^2)\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{V}(Y)+\mathbb{V}(X)]\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{E}(Y^2)+\mathbb{E}(X^2)]\}. \end{aligned} \tag{5.10}$$

In this case, σ_{XY} represents the difference between the sum of the variances of Y and X and the mean square difference $\mathbb{E}[(Y-X)^2]$.

In general, we thus have:

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{[\mathbb{V}(Y)+\mathbb{V}(X)] - \mathbb{V}(Y-X)\} \\ &= \frac{1}{2}\{\mathbb{V}(Y+X) - [\mathbb{V}(Y)+\mathbb{V}(X)]\}. \end{aligned} \tag{5.11}$$

If $\mu_Y = \mu_X$,

$$\sigma_{XY} = \frac{1}{2}\{[\mathbb{V}(Y)+\mathbb{V}(X)] - \mathbb{E}[(Y-X)^2]\} \tag{5.12}$$

and, if $\mu_Y = \mu_X = 0$,

$$\begin{aligned} \sigma_{XY} &= \frac{1}{2}\{[\mathbb{E}(Y^2)+\mathbb{E}(X^2)] - \mathbb{E}[(Y-X)^2]\} \\ &= \frac{1}{2}\{\mathbb{E}[(Y+X)^2] - [\mathbb{E}(Y^2)+\mathbb{E}(X^2)]\}. \end{aligned} \tag{5.13}$$

5.3. Correlations

From (5.5), it is also easy to see that

$$\mathbb{E}\left[\left(\frac{Y}{\sigma_Y} - \frac{X}{\sigma_X}\right)^2\right] = 2(1-\rho_{XY}) + \left(\frac{\mu_Y}{\sigma_Y} - \frac{\mu_X}{\sigma_X}\right)^2, \tag{5.14}$$

$$\mathbb{E}\left[\left(\frac{Y}{\sigma_Y} + \frac{X}{\sigma_X}\right)^2\right] = 2(1+\rho_{XY}) + \left(\frac{\mu_Y}{\sigma_Y} + \frac{\mu_X}{\sigma_X}\right)^2. \tag{5.15}$$

Consider the normalized values of X and Y :

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X}, \quad \tilde{Y} = \frac{Y - \mu_Y}{\sigma_Y}, \quad \rho(\tilde{X}, \tilde{Y}) = \rho\,(X, Y) := \rho_{XY}, \tag{5.16}$$

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

$$\mathbb{E}(\tilde{X}) = \mathbb{E}(\tilde{Y}) = 0, \quad \mathbb{V}(\tilde{X}) = \mathbb{V}(\tilde{Y}) = 1, \tag{5.17}$$

and

$$\mathbb{E}[(\tilde{Y} - \tilde{X})^2] = 2(1 - \rho_{XY}), \tag{5.18}$$

$$\rho_{XY} = 1 - \frac{1}{2}\mathbb{E}[(\tilde{Y} - \tilde{X})^2]. \tag{5.19}$$

The correlation $\rho\,(X, Y)$ is inversely related to the mean-square distance $\mathbb{E}[(\tilde{Y} - \tilde{X})^2]$ between $\tilde{X}$ and $\tilde{Y}$. (5.19) is a general form of the standard formula for Spearman's rank correlation coefficient.

Similarly,

$$\mathbb{E}[(\tilde{Y} + \tilde{X})^2] = 2(1 + \rho_{XY}), \tag{5.20}$$

$$\rho_{XY} = \frac{1}{2}\mathbb{E}[(\tilde{Y} + \tilde{X})^2] - 1. \tag{5.21}$$

The correlation $\rho\,(X, Y)$ measures the mean square $\mathbb{E}[(\tilde{Y} + \tilde{X})^2]$ of the sum of $\tilde{X} + \tilde{Y}$. The above formulae can also be rewritten in terms of the arithmetic mean of $\tilde{X}$ and $\tilde{Y}$:

$$\mathbb{E}\{\big[\frac{1}{2}(\tilde{Y} + \tilde{X})\big]^2\} = \frac{1}{2}(1 + \rho_{XY}), \tag{5.22}$$

$$\rho_{XY} = 2\mathbb{E}\{\big[\frac{1}{2}(\tilde{Y} + \tilde{X})\big]^2\} - 1 \tag{5.23}$$

5.4. Inequalities

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

$$(\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2 \leq \mathbb{E}[(Y - X)^2] \leq (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2, \tag{5.24}$$

and

$$\mathbb{E}[(Y - X)^2] \leq \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2 \leq (\sigma_Y + \sigma_X)^2 + (\mu_Y - \mu_X)^2, \text{ if } \rho_{XY} \geq 0, \tag{5.25}$$

$$\mathbb{E}[(Y - X)^2] \geq \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2 \geq (\sigma_Y - \sigma_X)^2 + (\mu_Y - \mu_X)^2, \text{ if } \rho_{XY} \leq 0, \tag{5.26}$$

$$\mathbb{E}[(Y - X)^2] = \sigma_Y^2 + \sigma_X^2 + (\mu_Y - \mu_X)^2, \text{ if } \rho_{XY} = 0. \tag{5.27}$$

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

$$\mathbb{E}[(Y-X)^2]=(\sigma_Y-\sigma_X)^2+(\mu_Y-\mu_X)^2\,,\quad \text{if } \rho_{XY}=1\,, \tag{5.28}$$

-

$$\mathbb{E}[(Y-X)^2]=(\sigma_Y+\sigma_X)^2+(\mu_Y-\mu_X)^2\,,\quad \text{if } \rho_{XY}=-1\,. \tag{5.29}$$

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

$$\left(1-\frac{\sigma_X}{\sigma_Y}\right)^2+\left(\frac{\mu_Y-\mu_X}{\sigma_Y}\right)^2\le\frac{\mathbb{E}[(Y-X)^2]}{\sigma_Y^2}\le\left(1+\frac{\sigma_X}{\sigma_Y}\right)^2+\left(\frac{\mu_Y-\mu_X}{\sigma_Y}\right)^2. \tag{5.30}$$

The inequalities (5.24) - (5.27) also entail similar properties for $X+Y$:

$$(\sigma_X-\sigma_Y)^2+(\mu_X+\mu_Y)^2\le\mathbb{E}[(X+Y)^2]\le(\sigma_X+\sigma_Y)^2+(\mu_X+\mu_Y)^2\,, \tag{5.31}$$

$$\mathbb{E}[(X+Y)^2]\le\sigma_X^2+\sigma_Y^2+(\mu_X+\mu_Y)^2\le(\sigma_Y+\sigma_X)^2+(\mu_X+\mu_Y)^2\,,\ \text{if } \rho_{XY}\le 0\,, \tag{5.32}$$

$$\mathbb{E}[(X+Y)^2]\ge\sigma_X^2+\sigma_Y^2+(\mu_X+\mu_Y)^2\ge(\sigma_X-\sigma_Y)^2+(\mu_X+\mu_Y)^2\,,\ \text{if } \rho_{XY}\ge 0\,, \tag{5.33}$$

$$\mathbb{E}[(Y+X)^2]=\sigma_X^2+\sigma_Y^2+(\mu_X+\mu_Y)^2\,,\ \text{if } \rho_{XY}=0\,. \tag{5.34}$$

By (5.18), we have:

$$0\le\mathbb{E}[(\tilde{Y}-\tilde{X})^2]\le 4\,, \tag{5.35}$$

$$0\le\mathbb{E}[|\tilde{Y}-\tilde{X}|]\le\{\mathbb{E}[(\tilde{Y}-\tilde{X})^2]\}^{1/2}\le 2\,. \tag{5.36}$$

The root mean square error of approximating $\tilde{Y}$ by $\tilde{X}$ cannot be larger than 2. Upon using the Chebyshev inequality, this entails:

$$\mathbb{P}\left[|\tilde{Y}-\tilde{X}|\ge\lambda\right]\le\frac{\mathbb{E}[(\tilde{Y}-\tilde{X})^2]}{\lambda^2}\le\frac{4}{\lambda^2}\,. \tag{5.37}$$

Since

$$X=\mu_X+\sigma_X\tilde{X}\,,\quad Y=\mu_Y+\sigma_Y\tilde{Y}\,, \tag{5.38}$$

we get

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

$$= \mathbb{E}[(\sigma_Y \tilde{Y} - \sigma_X \tilde{X})^2] + (\mu_Y - \mu_X)^2 \tag{5.39}$$

hence

$$\begin{aligned} \mathbb{E}[(Y-X)^2] &= \sigma_Y^2 \mathbb{E}\left[\left(\tilde{Y} - \frac{\sigma_X}{\sigma_Y}\tilde{X}\right)^2\right] + (\mu_Y - \mu_X)^2 \\ &= \sigma_Y^2 \left[1 + \left(\frac{\sigma_X}{\sigma_Y}\right)^2 - 2\left(\frac{\sigma_X}{\sigma_Y}\right)\rho_{XY}\right] + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y \neq 0, \end{aligned} \tag{5.40}$$

and

$$\mathbb{E}[(Y-X)^2] = \sigma_X^2 + (\mu_Y - \mu_X)^2, \quad \text{if } \sigma_Y = 0. \tag{5.41}$$

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

$$\sigma_Y^2 = \sigma_X^2, \tag{5.42}$$

we have:

$$\begin{aligned} \mathbb{E}[(Y-X)^2] &= 2\sigma_Y^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2 \\ &= 2\sigma_X^2(1-\rho_{XY}) + (\mu_Y - \mu_X)^2. \end{aligned} \tag{5.43}$$

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

$$\mu_Y = \mu_X \text{ and } \sigma_Y^2 = \sigma_X^2, \tag{5.44}$$

we have:

$$\mathbb{E}[(Y-X)^2] = 2\sigma_Y^2(1-\rho_{XY}) = 2\sigma_X^2(1-\rho_{XY}) \tag{5.45}$$

and

$$0 \leq \mathbb{E}[(Y-X)^2] \leq 4\sigma_X^2 \tag{5.46}$$

so that

$$\mathbb{E}[(Y-X)^2] = 0 \text{ and } \mathbb{P}[Y=X] = 1, \text{ if } \rho_{XY} = 1, \tag{5.47}$$

and, using Chebyshev's inequality,

$$\mathbb{P}[|Y-X| > c] \leq \frac{\mathbb{E}[(Y-X)^2]}{c^2} = \frac{2\sigma_X^2(1-\rho_{XY})}{c^2} \text{ for any } c > 0, \tag{5.48}$$

$$\mathbb{P}[|Y-X| > c\sigma_X] \leq \frac{\mathbb{E}[(Y-X)^2]}{\sigma_X^2 c^2} = \frac{2(1-\rho_{XY})}{c^2} \text{ for any } c > 0. \tag{5.49}$$

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

$$\mathbb{E}[(Y-X)^2] = 0 \Leftrightarrow \rho_{XY} = 1\,, \tag{5.50}$$

$$\mathbb{E}[(Y-X)^2] = 2\sigma_X^2 \Leftrightarrow \rho_{XY} = 0\,, \tag{5.51}$$

$$\mathbb{E}[(Y-X)^2] = 4\sigma_X^2 \Leftrightarrow \rho_{XY} = -1\,. \tag{5.52}$$

Since

$$\sigma_Y(\tilde{Y}-\tilde{X}) = Y - \mu_Y - \frac{\sigma_Y}{\sigma_X}(X-\mu_X) = Y - \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) - \frac{\sigma_Y}{\sigma_X}X\,, \tag{5.53}$$

the linear function

$$L_0(X) = \left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\mu_X\right) + \frac{\sigma_Y}{\sigma_X}X \tag{5.54}$$

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

$$\mathbb{E}[(Y-L_0(X))^2] = \sigma_Y^2\,\mathbb{E}[(\tilde{Y}-\tilde{X})^2] = 2\sigma_Y^2(1-\rho_{XY})\,. \tag{5.55}$$

It is then of interest to note that

$$\mathbb{E}[(Y-L_0(X))^2] \leq \mathbb{E}[(Y-\mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} \geq 0.5\,, \tag{5.56}$$

with

$$\mathbb{E}[(Y-L_0(X))^2] < \mathbb{E}[(Y-\mu_Y)^2] = \sigma_Y^2 \Leftrightarrow \rho_{XY} > 0.5 \tag{5.57}$$

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: $\mathbb{E}[(Y-L_0(X))^2] > \sigma_Y^2$.

5.5. Polarization identities

Since

$$\mathbb{E}[(Y-X)^2] = \mathbb{E}(X^2+Y^2-2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)-2\mathbb{E}(XY)\,, \tag{5.58}$$

$$\mathbb{E}[(Y+X)^2] = \mathbb{E}(X^2+Y^2+2XY) = \mathbb{E}(X^2)+\mathbb{E}(Y^2)+2\mathbb{E}(XY)\,, \tag{5.59}$$

we get on summing the above two equations:

$$\mathbb{E}(XY) = \frac{1}{4}\{\mathbb{E}[(Y+X)^2]-\mathbb{E}[(Y-X)^2]\}\,. \tag{5.60}$$

Similarly, since

$$\mathbb{V}(X+Y) = \mathbb{V}(X)+\mathbb{V}(Y)+2\mathsf{C}(X,Y)\,, \tag{5.61}$$

$$\mathbb{V}(X-Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathsf{C}(X,Y)\,, \tag{5.62}$$

we have:

$$\mathsf{C}(X,Y) = \frac{1}{4}[\mathbb{V}(X+Y) - \mathbb{V}(X-Y)]\,. \tag{5.63}$$

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

$$\rho\,(X,Y) = \frac{1}{4}\frac{\mathbb{V}(X+Y) - \mathbb{V}(X-Y)}{\sigma_X\,\sigma_Y} = \frac{1}{4}\left[\frac{\sigma^2_{X+Y}}{\sigma_X\,\sigma_Y} - \frac{\sigma^2_{X-Y}}{\sigma_X\,\sigma_Y}\right] \tag{5.64}$$

and, if $\mathbb{V}(X) = \mathbb{V}(Y) = 1$,

$$\rho\,(X,Y) = \frac{\mathbb{V}(X+Y) - \mathbb{V}(X-Y)}{4} = \frac{\sigma^2_{X+Y} - \sigma^2_{X-Y}}{4}\,. \tag{5.65}$$

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

$$\mathsf{C}(X+Y, X-Y) = [\mathbb{V}(X) - \mathbb{V}(Y)] + [\mathsf{C}(Y,X) - \mathsf{C}(X,Y)] = \mathbb{V}(X) - \mathbb{V}(Y) \tag{5.66}$$

so that

$$\mathsf{C}\big((X+Y)/2, X-Y\big) = \mathsf{C}(X+Y, X-Y) = 0\,, \quad \text{if } \mathbb{V}(X) = \mathbb{V}(Y)\,. \tag{5.67}$$

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 $\mathbb{V}(X) = \mathbb{V}(Y)$.

On applying (5.64) to the normalized variables $\tilde{Y}$ and $\tilde{X}$, we get a polarization formula in terms of normalized variables:

$$\rho\,(X,Y) = \frac{\mathbb{V}(\tilde{Y}+\tilde{X}) - \mathbb{V}(\tilde{Y}-\tilde{X})}{4} = \frac{\mathbb{E}[(\tilde{Y}+\tilde{X})^2] - \mathbb{E}[(\tilde{Y}-\tilde{X})^2]}{4}\,. \tag{5.68}$$

This also follows on applying (5.64) to $\tilde{Y}$ and $\tilde{X}$.

6. Hoeffding representation

The symbol $\int$ without lowerscript and upperscript represents integration over the whole real number set $(\mathbb{R})$.

Lemma 6.1 INDICATOR REPRESENTATION OF CONTINUOUS VARIABLE. *For any $x \in \mathbb{R}$,*

$$x = \int \{I[0 \le u \le x] - I[x \le u < 0]\}du$$

$$
\begin{aligned}
&= \int \{I[u \geq 0]I[u \leq x] - I[u < 0]I[u \geq x]\} du \\
&= \begin{cases} \int I[u \geq 0]I[u \leq x] du & \textit{for } x \geq 0 \\ -\int I[u < 0]I[u \geq x] du & \textit{for } x < 0 \end{cases} .
\end{aligned} \tag{6.1}
$$

For any x_1 , $x_2 \in \mathbb{R}$,

$$
\begin{aligned}
x_1 - x_2 &= \int \{I[x_2 \leq u] - I[x_1 \leq u]\} du \\
&= \int \{I[x_1 > u] - I[x_2 > u]\} du .
\end{aligned} \tag{6.2}
$$

PROOF OF LEMMA 6.1 To show (6.1), we use the fact that the functions $I[0 \leq u \leq x]$ and $I[x \leq u < 0]$ are indicator functions equal to zero everywhere except on finite intervals: for $x \geq 0$, we have $I[x \leq u < 0] = 0$ and

$$
\int \{I[0 \leq u \leq x] - I[x \leq u < 0]\} du = \int I[0 \leq u \leq x] du = \int_0^x du = x ; \tag{6.3}
$$

similarly, for $x < 0$, $I[0 \leq u \leq x] = 0$ and

$$
\int \{I[0 \leq u \leq x] - I[x \leq u < 0]\} du = -\int I[x \leq u < 0] du = -\int_x^0 du = x . \tag{6.4}
$$

On noting that

$$
I[0 \leq u \leq x] = I[u \geq 0]I[u \leq x] , \quad I[x \leq u < 0] = I[u < 0]I[u \geq x] , \tag{6.5}
$$

(6.1) follows. Setting

$$
\delta(x,u) := I[u \geq 0]I[u \leq x] - I[u < 0]I[u \geq x] , \tag{6.6}
$$

we can write:

$$
x_1 - x_2 = \int \{\delta(x_1,u) - \delta(x_2,u)\} du \tag{6.7}
$$

where

$$
\begin{aligned}
\delta(x_1,u) - \delta(x_2,u) &= \{I[u \geq 0]I[u \leq x_1] - I[u < 0]I[u \geq x_1]\} - \{I[u \geq 0]I[u \leq x_2] - I[u < 0]I[u \geq x_2]\} \\
&= I[u \geq 0]\{I[u \leq x_1] - I[u \leq x_2]\} - I[u < 0]\{I[u \geq x_1] - I[u \geq x_2]\} \\
&= I[u \geq 0]\{I[u \leq x_1] - I[u \leq x_2]\} - I[u < 0]\{(1 - I[u < x_1]) - (1 - I[u < x_2])\}
\end{aligned}
$$

$$\begin{aligned} &= I[u \geq 0]\{I[u \leq x_1] - I[u \leq x_2]\} + I[u < 0]\{I[u < x_2] - I[u < x_1]\} \\ &= \{I[u \leq x_1] - I[u \leq x_2]\} - I[u < 0]\{I[u = x_1] - I[u = x_2]\}\,. \end{aligned} \tag{6.8}$$

By the properties of the Riemann integral, we have

$$\int I[u < 0]\{I[u = x_1] - I[u = x_2]\}du = 0 \tag{6.9}$$

hence

$$x_1 - x_2 = \int \{\delta(x_1, u) - \delta(x_2, u)\}du \tag{6.10}$$

$$= \int \{I[x_2 \leq u] - I[x_1 \leq u]\}du - \int I[u < 0]\{I[u = x_1] - I[u = x_2]\}du \tag{6.11}$$

$$= \int \{I[x_2 \leq u] - I[x_1 \leq u]\}du \tag{6.12}$$

$$= \int \{(1 - I[x_2 > u]) - (1 - I[x_1 \leq u])\}du \tag{6.13}$$

$$= \int \{I[x_1 > u] - I[x_2 > u]\}du \tag{6.14}$$

and (6.2) is established. □

Theorem 6.2 HOEFFDING COVARIANCE IDENTITY. *Let $(X,Y)'$ be a pair of real random variables such that $\mathbb{P}[X \leq x, Y \leq y] := F(x,y)$, $\mathbb{P}[X \leq x] := F_X(x)$ and $\mathbb{P}[Y \leq y] := F_Y(y)$. If $(X,Y)'$ has finite second moments, then*

$$\begin{aligned} \mathsf{C}(X,Y) &= \int\int \big[F(x,y) - F_X(x)F_Y(y)\big]dx\,dy \\ &= \int\int \big[\mathbb{P}[X \leq x, Y \leq y] - \mathbb{P}[X \leq x]\mathbb{P}[Y \leq y]\big]dx\,dy \\ &= \int\int \big[\mathbb{P}[X > x, Y > y] - \mathbb{P}[X > x]\mathbb{P}[Y > y]\big]dx\,dy \\ &= \int\int \big[\mathsf{C}(I[X \leq x], I[Y \leq y])\big]dx\,dy \\ &= \int\int \big[\mathsf{C}(I[X > x], I[Y > y])\big]dx\,dy\,. \end{aligned} \tag{6.15}$$

PROOF OF THEOREM 6.2 Let $(X_1, Y_1)'$ and $(X_2, Y_2)'$ two i.i.d. vectors with joint distribution $F(x, y)$. Then

$$2\mathsf{C}(X, Y) = \mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)]\,. \tag{6.16}$$

By Lemma 6.1, we can write:

$$X_1 - X_2 = \int \{I[X_2 \le u] - I[X_1 \le u]\} du\,, \tag{6.17}$$

$$Y_1 - Y_2 = \int \{I[Y_2 \le v] - I[Y_1 \le v]\} dv\,, \tag{6.18}$$

where

$$\mathbb{E}(I[X_1 \le u]) = \mathbb{E}(I[X_2 \le u]) = F_X(u)\,, \tag{6.19}$$

$$\mathbb{E}(I[Y_1 \le v]) = \mathbb{E}(I[Y_2 \le v]) = F_Y(v)\,, \tag{6.20}$$

$$\mathbb{E}(I[X_1 \le u]I[Y_1 \le v]) = \mathbb{E}(I[X_2 \le u]I[Y_2 \le v]) = F(u, v)\,, \tag{6.21}$$

hence

$$(X_1 - X_2)(Y_1 - Y_2) = \int\int \{I[X_2 \le u] - I[X_1 \le u]\}\{I[Y_2 \le v] - I[Y_1 \le v]\} du\, dv\,. \tag{6.22}$$

Taking the expected value, we then get:

$$\begin{aligned}
\mathbb{E}[(X_1 - X_2)(Y_1 - Y_2)] &= \mathbb{E}\int\int \{I[X_2 \le u] - I[X_1 \le u]\}\{I[Y_2 \le v] - I[Y_1 \le v]\} du\, dv \\
&= \int\int \mathbb{E}[\{I[X_2 \le u] - I[X_1 \le u]\}\{I[Y_2 \le v] - I[Y_1 \le v]\}] du\, dv \\
&= 2\int\int [F(u, v) - F_X(u)F_Y(v)] du\, dv
\end{aligned} \tag{6.23}$$

where the expected value can be taken under the integral sign [by the fact that (X, Y) has finite second moments], and

$$\begin{aligned}
\mathsf{C}(X, Y) &= \int\int \big[F(x, y) - F_X(x)F_Y(y)\big] dx\, dy \\
&= \int\int \big[\mathbb{P}[X \le x, Y \le y] - \mathbb{P}[X \le x]\mathbb{P}[Y \le y]\big] dx\, dy\,.
\end{aligned} \tag{6.24}$$

The other identities in (6.15) follow from observing that

$$\mathsf{C}(I[X \le x], I[Y \le y]) = \mathbb{E}(I[X \le x]I[Y \le y]) - \mathbb{E}(I[X \le x])\mathbb{E}(I[Y \le y])$$

$$= \mathbb{P}[X \le x, Y \le y] - \mathbb{P}[X \le x]\mathbb{P}[Y \le y] = F(x,y) - F_X(x)F_Y(y)\,, \tag{6.25}$$

and

$$\begin{aligned}
\mathsf{C}(I[X &\le x], I[Y \le y]) = \mathsf{C}(1 - I[X > x], 1 - I[Y > y]) = \mathsf{C}(I[X > x], I[Y > y]) \\
&= \mathbb{E}(I[X > x]I[Y > y]) - \mathbb{E}(I[X > x])\mathbb{E}(I[Y > y]) \\
&= \mathbb{P}[X > x, Y > y] - \mathbb{P}[X > x]\mathbb{P}[Y > y]\,.
\end{aligned} \tag{6.26}$$

□

Corollary 6.3 HOEFFDING-TYPE VARIANCE REPRESENTATION. *Let X be a real random variable with distribution function $F_X(x)$ such that $\mathbb{E}(X^2) < \infty$.Then*

$$\begin{aligned}
\mathbb{V}(X) &= \int\int \big[\min\{F_X(x), F_X(y)\} - F_X(x)F_X(y)\big]dx\,dy \\
&= \int\int \big[\min\{F_X(x)[1 - F_X(y)], [1 - F_X(x)]F_X(y)\}\big]dx\,dy\,.
\end{aligned} \tag{6.27}$$

Corollary 6.4 HOEFFDING-TYPE CONDITIONAL DISTRIBUTION REPRESENTATION. *Let $(X,Y)'$ be a pair of real random variables with finite second moments. Then*

$$\begin{aligned}
\mathsf{C}(X,Y) &= \int\int \big[\{\mathbb{P}[Y \le y | X \le x] - \mathbb{P}[Y \le y]\}\mathbb{P}[X \le x]\big]dx\,dy \\
&= \int\int D(Y \le y | X \le x)\,\mathbb{P}[X \le x]dydx
\end{aligned} \tag{6.28}$$

where

$$D(Y \le y | X \le x) := \mathbb{P}[Y \le y | X \le x] - \mathbb{P}[Y \le y]\,. \tag{6.29}$$

Further, if the integral $\int D(Y \le y | X \le x)dy$ is finite for all x,

$$\mathsf{C}(X,Y) = \int \Big[\int D(Y \le y | X \le x)dy\Big]\mathbb{P}[X \le x]dx \tag{6.30}$$

Corollary 6.5 HOEFFDING COVARIANCE IDENTIT:Y LOCATION INVARIANCE. *Let $(X,Y)'$ be a pair of real random variables such that $\mathbb{P}[X \le x, Y \le y] := F(x,y)$, $\mathbb{P}[X \le x] := F_X(x)$ and $\mathbb{P}[Y \le y] := F_Y(y)$. If $(X,Y)'$ has finite second moments, then, for any a , $b \in \mathbb{R}$,*

$$\mathsf{C}(X,Y) = \int\int \big[F(x+a, y+b) - F_X(x+a)F_Y(y+b)\big]dx\,dy$$

$$
\begin{aligned}
&= \int\int \left[\mathbb{P}[X \le x+a, Y \le y+b] - \mathbb{P}[X \le x+a]\mathbb{P}[Y \le y+b]\right] dx\,dy \\
&= \int\int \left[\mathbb{P}[X > x+a, Y > y+b] - \mathbb{P}[X > x+a]\mathbb{P}[Y > y+b]\right] dx\,dy \\
&= \int\int \left[\mathsf{C}(I[X \le x+a], I[Y \le y+b])\right] dx\,dy \\
&= \int\int \left[\mathsf{C}(I[X > x+a], I[Y > y+b])\right] dx\,dy\,.
\end{aligned} \tag{6.31}
$$

Proposition 6.6 COPULA REPRESENTATION OF COVARIANCE. *Let $Z := (X,Y)'$ be a pair of real random variables such that $\mathbb{P}[X \le x, Y \le y] := F(x,y)$, $\mathbb{P}[X \le x] := F_X(x)$ and $\mathbb{P}[Y \le y] := F_Y(y)$. If $F(x,y)$ has the copula representation*

$$
F(x,y) = C_Z[F_X(x), F_Y(y)]\,, \tag{6.32}
$$

then

$$
\mathsf{C}(X,Y) = \int\int \{C_Z[F_X(x), F_Y(y)] - F_X(x)F_Y(y)\} dx\,dy\,. \tag{6.33}
$$

The Hoeffding representation provides a way of decomposing the covariance in terms of quantiles. This can be done by considering:

$$
\begin{aligned}
\mathsf{C}(X, I_1; Y, I_2) &:= \int_{I_2}\int_{I_1} \left[F(x,y) - F_X(x)F_Y(y)\right] dx\,dy \\
&= \int_{I_1}\int_{I_2} \{\mathbb{P}[Y \le y | X \le x] - \mathbb{P}[Y \le y]\}\mathbb{P}[X \le x]\big] dy\,dx
\end{aligned} \tag{6.34}
$$

where I_1 and I_2 are two subsets of $\mathbb{R}$. In particular, if $I_1 = \mathbb{R}$ and $I_2 = (-\infty, z]$, we have:

$$
\begin{aligned}
\mathsf{C}(X, I_1; Y, I_2) &:= \int \int_{-\infty}^{z} \left[F(x,y) - F_X(x)F_Y(y)\right] dy\,dx \\
&= \int \Big[\int_{-\infty}^{z} \big[\{\mathbb{P}[Y \le y | X \le x] - \mathbb{P}[Y \le y]\} dy\big] \mathbb{P}[X \le x]\Big]\, dx\,.
\end{aligned} \tag{6.35}
$$

Using (6.29), we can also define:

$$
\bar{\mathsf{C}}(X, Y; p_1, p_2) := \int_{-\infty}^{p_1} \Big[\int_{-\infty}^{p_2} D(Y \le y | X \le x)\Big] \mathbb{P}[X \le x]\, dx\,. \tag{6.36}
$$

7. Linear regression and correlations

In this section, we study the links between correlations and linear regressions as approximations between two variables. We first observe that the mean of a random variable X minimizes the distance between X and an arbitrary constant.

7.1. Linear approximation

Proposition 7.1 MEAN OPTIMALITY. *Let X be a random variable with finite second moment. Then, for any real constant a,*

$$\mathbb{E}[(X-\mu_X)^2] \leq \mathbb{E}[(X-a)^2] \tag{7.1}$$

and

$$\mathbb{E}[(X-\mu_X)^2] < \mathbb{E}[(X-a)^2] \quad \text{if } a \neq \mu_X\,. \tag{7.2}$$

Proposition 7.2 UNCENTERED REGRESSION OPTIMALITY. *Let (X,Y) be a pair of random variables with finite second moments, and set*

$$\begin{aligned} \bar{\beta} \quad &= \mathbb{E}(XY)/\mathbb{E}(X^2) \quad && \text{if } \mathbb{E}(X^2) > 0 \\ &= 0 && \text{otherwise}\,. \end{aligned} \tag{7.3}$$

Then,

$$\mathbb{E}[Y - \bar{\beta}X] = \mu_Y - \bar{\beta}\mu_X\,, \tag{7.4}$$

$$\mathbb{E}[X(Y-\bar{\beta}X)] = 0\,, \tag{7.5}$$

and, for any real constant b,

$$\mathbb{E}[(Y-\bar{\beta}X)^2] \leq \mathbb{E}[(Y-bX)^2]\,. \tag{7.6}$$

If $\mathbb{E}(X^2) > 0$, then

$$\mathbb{E}[X(Y-\bar{\beta}X)] = 0\,, \tag{7.7}$$

$$\mathbb{E}[(Y-\bar{\beta}X)^2] < \mathbb{E}[(Y-bX)^2] \quad \text{if } b \neq \bar{\beta}\,. \tag{7.8}$$

Let (X,Y) a pair of random variables with finite second moments, and set

$$\begin{aligned} U(Y \llcorner X) := U(Y \llcorner X) &:= (Y-\mu_Y) - \beta(X \lrcorner Y)(X-\mu_X) \\ &= Y - \beta(X \lrcorner Y)X - [\mu_Y - \beta(X \lrcorner Y)\mu_X] \end{aligned} \tag{7.9}$$

where $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$,

$$\beta(X \rightharpoondown Y) = \frac{\mathsf{C}(X,Y)}{\mathbb{V}(X)} = \frac{\rho\,(X,Y)\,\sigma(X)\,\sigma(Y)}{\mathbb{V}(X)} = \rho\,(X,Y)\,\frac{\sigma(Y)}{\sigma(X)}\,. \tag{7.10}$$

with $\beta(X \rightharpoondown Y) := 0$ when $\mathbb{V}(X) = 0$.

Proposition 7.3 CENTERED REGRESSION OPTIMALITY. *Let (X,Y) be a pair of random variables with finite second moments, and let $U(Y \llcorner X)$ be defined by (7.9). Then,*

$$\mathbb{E}[U(Y \llcorner X)] = 0\,, \tag{7.11}$$

$$\mathbb{E}[X\,U(Y \llcorner X)] = \mathsf{C}[X, U(Y \llcorner X)] = 0\,, \tag{7.12}$$

and, for any real constants a and b,

$$\mathbb{E}[U(Y \llcorner X)^2] \leq \mathbb{E}[(Y - a - bX)^2]\,. \tag{7.13}$$

If $\mathbb{V}(X) > 0$, then

$$\mathbb{E}[U(Y \llcorner X)^2] < \mathbb{E}[(Y - a - bX)^2] \quad \text{when } b \neq \beta(X \rightharpoondown Y) \text{ or } a \neq \mu_Y - \beta(X \rightharpoondown Y)\mu_X\,. \tag{7.14}$$

PROOF We have:

$$\begin{aligned} \mathbb{E}[U(Y \llcorner X)] &= \mathbb{E}[(Y - \mu_Y) - \beta(X \rightharpoondown Y)(X - \mu_X)] \\ &= \mathbb{E}(Y - \mu_Y) - \beta(X \rightharpoondown Y)\mathbb{E}(X - \mu_X) = 0\,, \end{aligned} \tag{7.15}$$

$$\begin{aligned} \mathbb{E}[X\,U(Y \llcorner X)] &= \mathsf{C}[X, U(Y \llcorner X)] \\ &= \mathsf{C}[X, \beta(X \rightharpoondown Y)X - \big(\mu_Y - \beta(X \rightharpoondown Y)\mu_X\big)] \\ &= \mathsf{C}[X, Y - \beta(X \rightharpoondown Y)X] \\ &= \mathsf{C}[X, Y] - \mathsf{C}[X, \beta(X \rightharpoondown Y)X] \\ &= \mathsf{C}[X, Y] - \beta(X \rightharpoondown Y)\mathsf{C}[X, X] \\ &= \mathsf{C}[X, Y] - \frac{\mathsf{C}(X,Y)}{\mathbb{V}(X)}\mathsf{C}[X, X] = 0\,. \end{aligned} \tag{7.16}$$

For any constant b_1, we have:

$$\begin{aligned} S(b_1) &:= \mathbb{E}\{[(Y - \mu_Y) - b_1(X - \mu_X)]^2\} \\ &= \mathbb{E}\{[\beta(X \rightharpoondown Y)(X - \mu_X) + U(Y \llcorner X) - b_1(X - \mu_X)]^2\} \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\{[(\beta(X \neg Y) - b_1)(X - \mu_X) + U(Y \llcorner X)]^2\} \\
&= (\beta(X \neg Y) - b)^2 \mathbb{E}[(X - \mu_X)^2] + \mathbb{E}[U(Y \llcorner X)^2] \\
&= (\beta(X \neg Y) - b)^2 \mathbb{V}(X) + \mathbb{E}[U(Y \llcorner X)^2] \\
&\geq \mathbb{E}[U(Y \llcorner X)^2] \geq 0 \qquad (7.17)
\end{aligned}
$$

with

$$
\mathbb{E}[(Y - bX)^2] > \mathbb{E}[U(Y \llcorner X)^2] \quad \text{if } \mathbb{V}(X) > 0 \text{ and } b \neq \beta(X \neg Y). \qquad (7.18)
$$

In other words, the value $b_1 = \beta(X \neg Y)$ minimizes $S(b_1)$; if $\mathbb{V}(X) > 0$, this minimum is unique. □

7.2. Regression coefficients as solutions of moment equations

The problem considered in Theorem 7.3 can also be interpreted as the solution of moment equations:

$$
\mathbb{E}\{1(Y - a - bX)\} = 0\,, \qquad (7.19)
$$

$$
\mathbb{E}\{X(Y - a - bX)\} = 0\,, \qquad (7.20)
$$

or, in matrix form,

$$
\mathbb{E}\left\{\begin{bmatrix} 1 \\ X \end{bmatrix}(Y - a - bX)\right\} = \mathbf{0}\,. \qquad (7.21)
$$

When $\mathbb{V}(X) > 0$, the solution to this problem is:

$$
b = \frac{\mathsf{C}(X,Y)}{\mathbb{V}(X)}\,, \quad a = \mu_Y - b\mu_X\,, \qquad (7.22)
$$

and it is unique. When $\mathbb{V}(X) = 0$, every value of b can be a solution with $a = \mu_Y - b\mu_X$.

7.3. Decompositions

$Y - \mu_Y$ is decomposed as the sum of orthogonal components:

$$
Y - \mu_Y = \beta(X \neg Y)(X - \mu_X) + U(Y \llcorner X) \qquad (7.23)
$$

where

$$
\begin{aligned}
\mathsf{C}[\beta(X \rightharpoondown Y)(X-\mu_X), U(Y \leftharpoondown X)] &= \mathsf{C}[\beta(X \rightharpoondown Y)X, U(Y \leftharpoondown X)] \\
&= \beta(X \rightharpoondown Y)\mathsf{C}[X, U(Y \leftharpoondown X)] = 0
\end{aligned}
\tag{7.24}
$$

so that $\beta(X \rightharpoondown Y)X$ is called the component part of Y "predicted" (or "explained") by X, while $U(Y \leftharpoondown X)$ is called the component part of Y "not predicted" (or "unexplained") by X. The interpretation may depend on the context. We also have:

$$
\begin{aligned}
\mathsf{C}[X, Y] &= \mathsf{C}[X, \beta(X \rightharpoondown Y)(X-\mu_X) + U(Y \leftharpoondown X)] \\
&= \beta(X \rightharpoondown Y)\mathbb{V}(X),
\end{aligned}
\tag{7.25}
$$

$$
\mathsf{C}[U(Y \leftharpoondown X), Y] = \mathsf{C}[U(Y \leftharpoondown X), U(Y \leftharpoondown X)] = \mathbb{V}[U(Y \leftharpoondown X)], \tag{7.26}
$$

$$
\begin{aligned}
\mathbb{V}(Y) &= \mathsf{C}[Y, Y] = \mathsf{C}[\beta(X \rightharpoondown Y)(X-\mu_X) + U(Y \leftharpoondown X), Y] \\
&= \beta(X \rightharpoondown Y)\mathsf{C}[X, Y] + \mathsf{C}[U(Y \leftharpoondown X), Y] \\
&= \beta(X \rightharpoondown Y)^2\mathbb{V}(X) + \mathbb{V}[U(Y \leftharpoondown X)] \\
&= \left[\frac{\mathsf{C}(X, Y)}{\mathbb{V}(X)}\right]^2 \mathbb{V}(X) + \mathbb{V}[U(Y \leftharpoondown X)] \\
&= \frac{\mathsf{C}(X, Y)^2}{\mathbb{V}(X)} + \mathbb{V}[U(Y \leftharpoondown X)] \\
&= \frac{\rho(X, Y)^2\mathbb{V}(X)\mathbb{V}(Y)}{\mathbb{V}(X)} + \mathbb{V}[U(Y \leftharpoondown X)] \\
&= \rho(X, Y)^2\mathbb{V}(Y) + \mathbb{V}[U(Y \leftharpoondown X)],
\end{aligned}
\tag{7.27}
$$

$$
\frac{\mathbb{V}(Y)}{\mathbb{V}(X)} = \beta(X \rightharpoondown Y)^2 + \frac{\mathbb{V}[U(Y \leftharpoondown X)]}{\mathbb{V}(X)} \geq \beta(X \rightharpoondown Y)^2. \tag{7.28}
$$

If $\mathbb{V}(Y) > 0$, we define the fraction of $\mathbb{V}(Y)$ predicted (or explained) by Y:

$$
R^2(Y \leftharpoondown X) := \frac{\mathbb{V}[\beta(X \rightharpoondown Y)X]}{\mathbb{V}(Y)}. \tag{7.29}
$$

$R^2(Y \llcorner X)$ is called the *coefficient of determination* of Y on X (or from X to Y). We have:

$$R^2(Y \llcorner X) = \frac{\mathbb{V}[\beta(X \lrcorner Y)X]}{\mathbb{V}(Y)} = \rho(X,Y)^2\,, \tag{7.30}$$

$$\rho(X,Y)^2 + \frac{\mathbb{V}[U(Y \llcorner X)]}{\mathbb{V}(Y)} = 1\,, \tag{7.31}$$

$$\rho(X,Y)^2 = 1 - \frac{\mathbb{V}[U(Y \llcorner X)]}{\mathbb{V}(Y)}\,, \tag{7.32}$$

$$\frac{\mathbb{V}[U(Y \llcorner X)]}{\mathbb{V}(Y)} = 1 - \rho(X,Y)^2, \tag{7.33}$$

$$\rho(X,Y)^2 = 1 \Leftrightarrow \mathbb{V}[U(Y \llcorner X)] = 0\,. \tag{7.34}$$

7.4. Population t and F coefficients

If $\mathbb{V}[U(Y \llcorner X)] > 0$, then $\mathbb{V}(Y) > 0$ and the above identities can also be formulated in terms of F-type and t-type variables:

$$\begin{aligned}
\mathscr{F}(X \lrcorner Y) &:= \frac{\mathbb{V}(Y) - \mathbb{V}[U(Y \llcorner X)]}{\mathbb{V}[U(Y \llcorner X)]} \\
&= \frac{\mathbb{V}[\beta(X \lrcorner Y)X]}{\mathbb{V}[U(Y \llcorner X)]} = \frac{\beta(X \lrcorner Y)^2\,\mathbb{V}(X)}{\mathbb{V}[U(Y \llcorner X)]} \\
&= \left[\frac{\beta(X \lrcorner Y)\,\sigma(X)}{\sigma[U(Y \llcorner X)]}\right]^2 = t(X \lrcorner Y)^2
\end{aligned} \tag{7.35}$$

where

$$\begin{aligned}
t(X \lrcorner Y) &:= \frac{\beta(X \lrcorner Y)\,\sigma(X)}{\sigma[U(Y \llcorner X)]} = \frac{\sigma(X)}{\sigma[U(Y \llcorner X)]}\,\beta(X \lrcorner Y) \\
&= \frac{\beta(X \lrcorner Y)}{\{\mathbb{V}[U(Y \llcorner X)]\,\mathbb{V}(X)^{-1}\}^{1/2}}\,.
\end{aligned} \tag{7.36}$$

While $\beta(X \lrcorner Y)$ is a population linear regression coefficient, $\mathbb{V}[U(Y \llcorner X)]\,\mathbb{V}(X)^{-1}$ can be interpreted as the population analogue of the corresponding "standard error".

If $\rho(X,Y)^2 \neq 1$, we can also write:

$$
\begin{aligned}
\mathscr{F}(X \rightharpoondown Y) &= \frac{\mathbb{V}(Y)}{\mathbb{V}[U(Y \leftharpoondown X)]} - 1 = \frac{1}{1-\rho(X,Y)^2} - 1 \\
&= \frac{\rho(X,Y)^2}{1-\rho(X,Y)^2} \\
&= \frac{R^2(Y \leftharpoondown X)}{1-R^2(Y \leftharpoondown X)}, \qquad (7.37)
\end{aligned}
$$

$$
t(X \rightharpoondown Y) = \frac{\rho(X,Y)}{[1-\rho(X,Y)^2]^{1/2}}, \qquad (7.38)
$$

$$
\rho(X,Y)^2 = \frac{\mathscr{F}(X \rightharpoondown Y)}{1+\mathscr{F}(X \rightharpoondown Y)}, \qquad (7.39)
$$

$$
\rho(X,Y) = \frac{t(X \rightharpoondown Y)}{[1+t(X \rightharpoondown Y)^2]^{1/2}}, \qquad (7.40)
$$

$$
\beta(X \rightharpoondown Y) = \rho\,(X,Y)\,\frac{\sigma(Y)}{\sigma(X)} = \frac{t(X \rightharpoondown Y)}{[1+t(X \rightharpoondown Y)^2]^{1/2}}\frac{\sigma(Y)}{\sigma(X)}. \qquad (7.41)
$$

Since $\rho(Y,X) = \rho(X,Y)$, this entails that $\mathscr{F}(X \rightharpoondown Y)$ and $t(X \rightharpoondown Y)$ enjoy a symmetry property:

$$
\mathscr{F}(Y \rightharpoondown X) = \mathscr{F}(X \rightharpoondown Y), \quad t(Y \rightharpoondown X) = t(X \rightharpoondown Y). \qquad (7.42)
$$

However, symmetry does not hold for $\beta(Y \rightharpoondown X)$:

$$
\begin{aligned}
\beta(Y \rightharpoondown X) &= \rho\,(Y,X)\,\frac{\sigma(X)}{\sigma(Y)} \\
&= \rho\,(X,Y)\,\frac{\sigma(Y)}{\sigma(X)}\frac{\sigma(X)^2}{\sigma(Y)^2} \\
&= \beta(X \rightharpoondown Y)\frac{\sigma(X)^2}{\sigma(Y)^2} \qquad (7.43)
\end{aligned}
$$

except when $\sigma(Y) = \sigma(X)$ or $\rho(X, Y) = 0$. If $\rho(Y, X) \neq 0$, $\beta(Y \dashv X)$ and $\beta(X \dashv Y)$ have the same sign, with

$$\frac{\beta(Y \dashv X)}{\beta(X \dashv Y)} = \frac{\sigma(X)^2}{\sigma(Y)^2}, \tag{7.44}$$

$$\frac{\beta(Y \dashv X)}{\beta(X \dashv Y)} > 1 \Leftrightarrow \sigma(X) > \sigma(Y), \tag{7.45}$$

and the difference becomes larger as the ratio $\sigma(X)/\sigma(Y)$ increases. Instead, $\beta(X \dashv Y)$ satisfies the following weighted symmetry properties:

$$\frac{\beta(X \dashv Y)}{\sigma(Y)^2} = \frac{\beta(Y \dashv X)}{\sigma(X)^2}, \tag{7.46}$$

$$\beta\left(\frac{X}{\sigma(X)} \dashv \frac{Y}{\sigma(Y)}\right) = \beta\left(\frac{Y}{\sigma(Y)} \dashv \frac{X}{\sigma(X)}\right). \tag{7.47}$$

If each variable is divided by its standard error, the regression coefficient is the same irrespective of the variable selected as "dependent variable". Further,

$$0 \leq \beta(X \dashv Y)\beta(Y \dashv X) = \frac{\mathsf{C}(X, Y)}{\sigma(X)^2}\frac{\mathsf{C}(X, Y)}{\sigma(Y)^2} = \rho(X, Y)^2 \leq 1. \tag{7.48}$$

7.5. Inequalities on linear regression coefficients

From (7.10), we get the following inequality: since $|\rho(X, Y)| \leq 1$,

$$-\frac{\sigma(Y)}{\sigma(X)} \leq \beta(X \dashv Y) \leq \frac{\sigma(Y)}{\sigma(X)} \tag{7.49}$$

if $\sigma(X) > 0$, and

$$-\frac{\sigma(X)}{\sigma(Y)} \leq \beta(Y \dashv X) \leq \frac{\sigma(X)}{\sigma(Y)} \tag{7.50}$$

if $\sigma(Y) > 0$. Using (7.41), we also have: for $\sigma(Y) > 0$ and $\sigma(X) > 0$,

$$\begin{array}{ll} \beta(X \dashv Y) \leq \left(t_0/[1+t_0^2]^{1/2}\right)\frac{\sigma(Y)}{\sigma(X)} < 0 & \text{if } t(X \dashv Y) \leq t_0 < 0 \\ \left(t_0/[1+t_0^2]^{1/2}\right)\frac{\sigma(Y)}{\sigma(X)} < \beta(X \dashv Y) < 0 & \text{if } t_0 < t(X \dashv Y) < 0 \\ 0 \leq \beta(X \dashv Y) \leq \left(t_0/[1+t_0^2]^{1/2}\right)\frac{\sigma(Y)}{\sigma(X)} & \text{if } 0 \leq t(X \dashv Y) \leq t_0 \\ \beta(X \dashv Y) > \left(t_0/[1+t_0^2]^{1/2}\right)\frac{\sigma(Y)}{\sigma(X)} > 0 & \text{if } t(X \dashv Y) > t_0 > 0. \end{array} \tag{7.51}$$

8. Covariance and variance decompositions

We study here covariance and variance decompositions for sums of random variables. We assume that all the variables considered have finite second moments.

We consider in turn the following cases:

$$M := \sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \cdots + Y_n \,, \tag{8.1}$$

$$Y = M + U = \sum_{i=1}^{n} Y_i + U \,, \quad \mathsf{C}(M, U) = 0 \,, \tag{8.2}$$

$$M(\lambda) := \sum_{i=1}^{n} \lambda_i X_i = \lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_n X_n \,, \tag{8.3}$$

$$Y = M(\lambda) + U = \sum_{i=1}^{n} \lambda_i X_i + U \,, \quad \mathsf{C}(M(\lambda), U) = 0 \,, \tag{8.4}$$

8.1. Sum of random variables

8.1.1. Covariance decomposition

Consider the following sum of random variables:

$$M := \sum_{i=1}^{n} Y_i = Y_1 + Y_2 + \cdots + Y_n \,. \tag{8.5}$$

For any random variable Z, we have:

$$\mathsf{C}(Z, M) = \mathsf{C}\Big(Z, \sum_{i=1}^{n} Y_i\Big) = \sum_{i=1}^{n} \mathsf{C}(Z, Y_i) \,, \tag{8.6}$$

$$\beta(Z \dashv M) := \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \frac{\mathsf{C}(Z, Y_i)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \beta(Z \dashv Y_i) \,, \tag{8.7}$$

$$\beta(Z \dashv M) Z = \sum_{i=1}^{n} \beta(Z \dashv Y_i) Z \,, \tag{8.8}$$

where we set $\beta(Z \dashv M) := 0$ and $\beta(Z \dashv Y_i) := 0$ when $\mathbb{V}(Z) = 0$. $\mathsf{C}(Z, Y_i)$ can be interpreted as the contribution of Y_i to the covariance $\mathsf{C}(Z, M)$, while $\beta(Z \dashv Y_i)$ is the corresponding contribution as a proportion of the variance $\mathbb{V}(Z)$. These contributions can be positive or negative. Each component only depends on one pair (Z, Y_i), not on Y_j for $j \neq i$.

When Z and M are interchanged, we get:

$$
\begin{aligned}
\beta(M \rightharpoondown Z) &= \frac{\mathsf{C}(M, Z)}{\mathbb{V}(M)} = \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \beta(Z \rightharpoondown M) \\
&= \sum_{i=1}^{n} \frac{\mathsf{C}(Y_i, Z)}{\mathbb{V}(M)} = \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \beta(Z \rightharpoondown Y_i)\,, \qquad (8.9)
\end{aligned}
$$

$$
\beta(M \rightharpoondown Z)Z = \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \beta(Z \rightharpoondown M)Z = \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \beta(Z \rightharpoondown Y_i)Z\,, \qquad (8.10)
$$

where

$$
\beta(Z \rightharpoondown Y_i) := \frac{\mathsf{C}(Z, Y_i)}{\mathbb{V}(Z)}\,, \quad i = 1, \dots, n\,. \qquad (8.11)
$$

Set

$$
U(M \llcorner Z) := M - \beta(Z \rightharpoondown M)Z\,, \qquad (8.12)
$$

$$
U(Y_i \llcorner Z) := Y_i - \beta(Z \rightharpoondown Y_i)Z\,, \quad i = 1, \dots, n\,. \qquad (8.13)
$$

Then

$$
\begin{aligned}
\mathsf{C}[Z, U(M \llcorner Z)] &= \mathsf{C}[Z, M] - \mathsf{C}[Z, \beta(Z \rightharpoondown M)Z] \\
&= \mathsf{C}[Z, M] - \beta(Z \rightharpoondown M)\mathsf{C}[Z, Z] \\
&= \mathsf{C}[Z, M] - \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} \mathsf{C}[Z, Z] = 0\,, \qquad (8.14)
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{C}[Z, U(Y_i \llcorner Z)] &= \mathsf{C}[Z, Y_i] - \mathsf{C}[Z, \beta(Z \rightharpoondown Y_i)Z] \\
&= \mathsf{C}[Z, Y_i] - \frac{\mathsf{C}(Z, Y_i)}{\mathbb{V}(Z)} \mathsf{C}[Z, Z] = 0\,, \quad i = 1, \dots, n\,, \qquad (8.15)
\end{aligned}
$$

and we can write:

$$
M = \beta(Z \rightharpoondown M)Z + U(M \llcorner Z)\,, \quad \mathsf{C}[Z, U(M \llcorner Z)] = \mathsf{C}[\beta(Z \rightharpoondown M)Z, U(M \llcorner Z)] = 0\,, \qquad (8.16)
$$

$$
Y_i = \beta(Z \rightharpoondown Y_i)Z + U(Y_i \llcorner Z)\,, \quad \mathsf{C}[Z, U(Y_i \llcorner Z)] = \mathsf{C}[\beta(Z \rightharpoondown Y_i)Z, U(Y_i \llcorner Z)] = 0\,, \quad i = 1, \dots, n\,, \qquad (8.17)
$$

$$
\mathbb{V}(M) = \mathbb{V}[\beta(Z \rightharpoondown M)Z] + \mathbb{V}[U(M \llcorner Z)]
$$

$$
\begin{aligned}
&= \beta(Z \rightharpoondown M)^2 \mathbb{V}(Z) + \mathbb{V}[U(M \llcorner Z)] \\
&= \frac{\mathsf{C}(Z,M)^2}{\mathbb{V}(Z)} + \mathbb{V}[U(M \llcorner Z)] = \frac{\rho(Z,M)^2 \mathbb{V}(Z)\mathbb{V}(M)}{\mathbb{V}(Z)} + \mathbb{V}[U(M \llcorner Z)] \\
&= \rho(Z,M)^2 \mathbb{V}(M) + \mathbb{V}[U(M \llcorner Z)]\,,
\end{aligned}
\tag{8.18}
$$

$$
\rho(Z,M)^2 = 1 - \frac{\mathbb{V}[U(M \llcorner Z)]}{\mathbb{V}(M)}\,, \quad \rho(M,Z)^2 = 1 - \frac{\mathbb{V}[U(Z \llcorner M)]}{\mathbb{V}(Z)}\,,
\tag{8.19}
$$

$$
\frac{\mathbb{V}[U(M \llcorner Z)]}{\mathbb{V}(M)} = 1 - \rho(Z,M)^2 = 1 - \rho(M,Z)^2 = \frac{\mathbb{V}[U(Z \llcorner M)]}{\mathbb{V}(Z)}\,,
\tag{8.20}
$$

$$
\begin{aligned}
\mathbb{V}(Y_i) &= \mathbb{V}[\beta(Z \rightharpoondown Y_i)Z] + \mathbb{V}[U(Y_i \llcorner Z)] \\
&= \beta(Z \rightharpoondown Y_i)^2 \mathbb{V}(Z) + \mathbb{V}[U(Y_i \llcorner Z)] \\
&= \frac{\mathsf{C}(Z,Y_i)^2}{\mathbb{V}[U(M \llcorner Z)]} + \mathbb{V}[U(Y_i \llcorner Z)] \\
&= \rho(Z,Y_i)^2 \mathbb{V}(Y_i) + \mathbb{V}[U(Y_i \llcorner Z)]\,,
\end{aligned}
\tag{8.21}
$$

$$
\rho(Z,Y_i)^2 = 1 - \frac{\mathbb{V}[U(Y_i \llcorner Z)]}{\mathbb{V}(Y_i)}\,, \quad \rho(Y_i,Z)^2 = 1 - \frac{\mathbb{V}[U(Z \llcorner Y_i)]}{\mathbb{V}(Z)}\,,
\tag{8.22}
$$

$$
\frac{\mathbb{V}[U(Y_i \llcorner Z)]}{\mathbb{V}(Y_i)} = 1 - \rho(Z,Y_i)^2 = 1 - \rho(Y_i,Z)^2 = \frac{\mathbb{V}[U(Z \llcorner Y_i)]}{\mathbb{V}(Z)}\,,
\tag{8.23}
$$

$$
\mathsf{C}(Z,M) = \mathsf{C}[Z, \beta(Z \rightharpoondown M)Z + U(Z \llcorner M)] = \mathsf{C}[Z, \beta(Z \rightharpoondown M)Z]\,,
\tag{8.24}
$$

$$
\mathsf{C}(Z,Y_i) = \mathsf{C}[Z, \beta(Z \rightharpoondown Y_i)Z + U(Z \llcorner Y_i)] = \mathsf{C}[Z, \beta(Z \rightharpoondown Y_i)Z]\,.
\tag{8.25}
$$

$\beta(Z \rightharpoondown M)Z$ is the part of Z which contributes to $\mathsf{C}(Z,M)$, while $\beta(Z \rightharpoondown Y_i)Z$ is the part of Z which contributes to $\mathsf{C}(Z,Y_i)$.

The above identities can also be formulated in terms of F-type and t-type variables. Suppose that $\mathbb{V}[U(Z \llcorner M)] > 0$ and $\mathbb{V}(Y_i) > 0$, $i = 1, \ldots, n$. We can write:

$$
\begin{aligned}
\mathscr{F}(Z \rightharpoondown M) \quad :&= \frac{\mathbb{V}(M) - \mathbb{V}[U(M \llcorner Z)]}{\mathbb{V}[U(M \llcorner Z)]} \\
&= \frac{\mathbb{V}[\beta(Z \rightharpoondown M)Z]}{\mathbb{V}[U(M \llcorner Z)]} = \frac{\beta(Z \rightharpoondown M)^2\, \mathbb{V}(Z)}{\mathbb{V}[U(M \llcorner Z)]} \\
&= \left[\frac{\beta(Z \rightharpoondown M)\,\sigma(Z)}{\sigma[U(M \llcorner Z)]}\right]^2 = t(M,Z)^2
\end{aligned}
\tag{8.26}
$$

$$
\begin{aligned}
\mathscr{F}(Z \rightharpoonup M) &= \frac{\mathbb{V}(M)}{\mathbb{V}[U(M \llcorner Z)]} - 1 \\
&= \frac{1}{1-\rho(Z,M)^2} - 1 = \frac{\rho(Z,M)^2}{1-\rho(Z,M)^2} \qquad (8.27)
\end{aligned}
$$

where

$$
\begin{aligned}
t(Z \rightharpoonup M) \quad : &= \frac{\beta(Z \rightharpoonup M)\,\sigma(Z)}{\sigma[U(M \llcorner Z)]} = \frac{\sigma(Z)}{\sigma[U(M \llcorner Z)]}\,\beta(Z \rightharpoonup M) \\
&= \frac{\beta(Z \rightharpoonup M)}{\{\mathbb{V}[U(M \llcorner Z)]\,\mathbb{V}(Z)^{-1}\}^{1/2}} \\
&= \frac{\rho(Z,M)}{[1-\rho(Z,M)^2]^{1/2}}\,. \qquad (8.28)
\end{aligned}
$$

$\mathscr{F}(Z \rightharpoonup M)$ can be interpreted as the theoretical F-ratio associated with the regression of M on Z, while $t(Z \rightharpoonup M)$ can be interpreted the corresponding theoretical t-ratio. Further,

$$
\mathscr{F}(Z \rightharpoonup M) = \frac{\mathbb{V}(M)}{\mathbb{V}[U(M \llcorner Z)]} - 1 \qquad (8.29)
$$

$$
\begin{aligned}
t(Z \rightharpoonup M) &= \frac{\sigma(Z)}{\sigma[U(Z \llcorner M)]}\frac{\mathsf{C}(M,Z)}{\mathbb{V}(Z)} \\
&= \frac{\sigma(Z)}{\sigma[U(M \llcorner Z)]}\frac{\rho(M,Z)\sigma(M)\sigma(Z)}{\sigma(Z)^2} \\
&= \rho(M,Z)\frac{\sigma(M)}{\sigma[U(M \llcorner Z)]} \\
&= \rho(Z,M)\frac{\sigma(Z)}{\sigma[U(Z \llcorner M)]} \\
&= t(M,Z) \qquad (8.30)
\end{aligned}
$$

Similarly,

$$
\mathscr{F}(M,Z) \quad : \quad = \frac{\mathbb{V}(Z) - \mathbb{V}[U(Z \llcorner M)]}{\mathbb{V}[U(Z \llcorner M)]}
$$

$$= \quad \left[\frac{\beta(M \rightharpoonup Z)\,\sigma(M)}{\sigma[U(Z \leftharpoondown M)]}\right]^2 = t(M,Z)^2\,, \tag{8.31}$$

$$\begin{aligned} t(M,Z) \quad &: \quad = \frac{\sigma(M)}{\sigma[U(Z \leftharpoondown M)]}\beta(M \rightharpoonup Z) \\ &= \quad \frac{\beta(M \rightharpoonup Z)}{\{\mathbb{V}[U(Z \leftharpoondown M)]\mathbb{V}(M)^{-1}\}^{1/2}}\,, \end{aligned} \tag{8.32}$$

$$\begin{aligned} \mathscr{F}(Z,Y_i) \quad &: \quad = \frac{\mathbb{V}(Y_i) - \mathbb{V}[U(Y_i \leftharpoondown Z)]}{\mathbb{V}[U(Y_i \leftharpoondown Z)]} \\ &= \quad \frac{\mathbb{V}[\beta(Z \rightharpoonup Y_i)Z]}{\mathbb{V}[U(Y_i \leftharpoondown Z)]} = \frac{\beta(Z \rightharpoonup Y_i)^2\,\mathbb{V}(Z)}{\mathbb{V}[U(Y_i \leftharpoondown Z)]} \\ &= \quad \left[\frac{\beta(Z \rightharpoonup Y_i)\,\sigma(Z)}{\sigma[U(Y_i \leftharpoondown Z)]}\right]^2 = t(Z,Y_i)^2 \end{aligned} \tag{8.33}$$

$$\begin{aligned} t(Z,Y_i) \quad &: \quad = \frac{\beta(Z \rightharpoonup Y_i)\,\sigma(Z)}{\sigma[U(Y_i \leftharpoondown Z)]} = \frac{\sigma(Z)}{\sigma[U(Y_i \leftharpoondown Z)]}\beta(Z \rightharpoonup Y_i) \\ &= \quad \frac{\beta(Z \rightharpoonup Y_i)}{\{\mathbb{V}[U(Y_i \leftharpoondown Z)]\mathbb{V}(Z)^{-1}\}^{1/2}}\,. \end{aligned} \tag{8.34}$$

$$\begin{aligned} \mathscr{F}(Y_i,Z) \quad &: \quad = \frac{\mathbb{V}(Z) - \mathbb{V}[U(Z \leftharpoondown Y_i)]}{\mathbb{V}[U(Z \leftharpoondown Y_i)]} \\ &= \quad \left[\frac{\beta(Y_i,Z)\sigma(Y_i)}{\sigma[U(Z \leftharpoondown Y_i)]}\right]^2 = t(Y_i,Z)^2\,, \end{aligned} \tag{8.35}$$

$$\begin{aligned} t(Y_i,Z) \quad &: \quad = \frac{\sigma(Y_i)}{\sigma[U(Z \leftharpoondown Y_i)]}\beta(Y_i,Z) \\ &= \quad \frac{\beta(Y_i,Z)}{\{\mathbb{V}[U(Z \leftharpoondown Y_i)]\mathbb{V}(Y_i)^{-1}\}^{1/2}}\,. \end{aligned} \tag{8.36}$$

We thus have the following decompositions:

$$\begin{aligned}
\beta(M \rightharpoonup Z) &= \frac{\sigma[U(Z \leftharpoonup M)]}{\sigma(M)} t(M, Z) \\
&= \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \beta(Z \rightharpoonup Y_i) \\
&= \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \frac{\sigma[U(Y_i \leftharpoonup Z)]}{\sigma(Z)} t(Z, Y_i) \\
&= \frac{\sigma(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \sigma[U(Y_i \leftharpoonup Z)]\, t(Y_i, Z) \qquad (8.37)
\end{aligned}$$

$$\begin{aligned}
\beta(Z \rightharpoonup M) \quad &: \quad = \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \beta(Z \rightharpoonup Y_i) \\
&= \sum_{i=1}^{n} \frac{\sigma[U(Y_i \leftharpoonup Z)]}{\sigma(Z)} t(Z, Y_i) = \frac{1}{\sigma(Z)} \sum_{i=1}^{n} \sigma[U(Y_i \leftharpoonup Z)]\, t(Z, Y_i)\,, \qquad (8.38)
\end{aligned}$$

hence

$$\begin{aligned}
t(M, Z) &= \frac{\sigma(M)}{\sigma[U(Z \leftharpoonup M)]} \beta(M \rightharpoonup Z) \\
&= \frac{\sigma(Z)}{\sigma[U(Z \leftharpoonup M)]\sigma(M)} \sum_{i=1}^{n} \sigma[U(Y_i \leftharpoonup Z)]\, t(Y_i, Z) \\
&= \frac{\sigma(Z)}{\sigma(M)} \sum_{i=1}^{n} \frac{\sigma[U(Y_i \leftharpoonup Z)]}{\sigma[U(Z \leftharpoonup M)]} t(Y_i, Z)\,, \qquad (8.39)
\end{aligned}$$

$$\begin{aligned}
t(Z \rightharpoonup M) \quad &: \quad = \frac{\sigma(Z)}{\sigma[U(M \leftharpoonup Z)]} \beta(Z \rightharpoonup M) \\
&= \frac{\sigma(Z)}{\sigma[U(M \leftharpoonup Z)]} \sum_{i=1}^{n} \beta(Z \rightharpoonup Y_i) \\
&= \frac{\sigma(Z)}{\sigma[U(M \leftharpoonup Z)]} \sum_{i=1}^{n} \frac{\sigma[U(Y_i \leftharpoonup Z)]}{\sigma(Z)} t(Z, Y_i)
\end{aligned}$$

$$= \sum_{i=1}^{n} \frac{\sigma[U(Y_i \llcorner Z)]}{\sigma[U(M \llcorner Z)]} t(Z, Y_i) \,. \tag{8.40}$$

Note also that:

$$\begin{aligned} t(M, Z) &= \frac{\sigma(M)}{\sigma[U(Z \llcorner M)]} \beta(M \lrcorner Z) \\ &= \frac{\sigma(M)}{\sigma[U(Z \llcorner M)]} \frac{\mathsf{C}(M, Z)}{\mathbb{V}(M)} \\ &= \frac{\sigma(M)}{\sigma[U(Z \llcorner M)]} \frac{\rho(M, Z)\sigma(M)\sigma(Z)}{\mathbb{V}(M)} \\ &= \frac{\sigma(Z)}{\sigma[U(Z \llcorner M)]} \rho(M, Z) \,. \end{aligned} \tag{8.41}$$

8.1.2. Covariance-variance decomposition

Consider now the case where

$$Z = M \quad \text{and} \quad \mathbb{V}(M) > 0 \tag{8.42}$$

and set

$$\beta(M \lrcorner Y_i) := \frac{\mathsf{C}(M, Y_i)}{\mathbb{V}(M)} \,, \quad i = 1, \ldots, n \,. \tag{8.43}$$

We then have:

$$\mathbb{V}(M) = \mathsf{C}[M, \sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} \mathsf{C}(M, Y_i) \,, \tag{8.44}$$

$$\sum_{i=1}^{n} \frac{\mathsf{C}(M, Y_i)}{\mathbb{V}(M)} = \sum_{i=1}^{n} \beta(M \lrcorner Y_i) = 1 \,, \tag{8.45}$$

$$\beta(M \lrcorner Y_i) := \frac{\mathsf{C}(M, Y_i)}{\mathbb{V}(Y_i)} \frac{\mathbb{V}(Y_i)}{\mathbb{V}(M)} = \beta(Y_i, M) \frac{\mathbb{V}(Y_i)}{\mathbb{V}(M)} \,. \tag{8.46}$$

$\mathsf{C}(M, Y_i)$ can be interpreted as the contribution of Y_i to the variance $\mathbb{V}(M)$, while $\beta(M \lrcorner Y_i)$ is the corresponding contribution as a proportion of $\mathbb{V}(M)$. These contributions can be positive or negative. Then

$$\beta(Z \lrcorner M) = \beta(M \lrcorner M) = 1 \,, \tag{8.47}$$

$$U(Z \llcorner M) = U(M \llcorner M) = M - \beta(M \lrcorner M) M = 0 \,, \tag{8.48}$$

$$U(Y_i \llcorner M) = Y_i - \beta(M \lrcorner Y_i) M\,, \quad i = 1, \ldots, n\,, \tag{8.49}$$

$$\mathsf{C}[M, U(M \llcorner M)] = 0\,, \tag{8.50}$$

$$\mathsf{C}[M, U(Y_i \llcorner M)] = 0\,, \quad i = 1, \ldots, n\,, \tag{8.51}$$

and we can write:

$$M = \beta(Z \lrcorner M) M = \beta(M \lrcorner M) M\,, \tag{8.52}$$

$$\mathsf{C}(Z, M) = \mathsf{C}(M, M) = \mathbb{V}(M)\,, \tag{8.53}$$

$$Y_i = \beta(M \lrcorner Y_i) Z + U(Y_i \llcorner M)\,, \quad \mathsf{C}[M, U(Y_i \llcorner M)] = 0\,, \quad i = 1, \ldots, n\,, \tag{8.54}$$

$$\mathsf{C}(M, Y_i) = \mathsf{C}[M, \beta(M \lrcorner Y_i) Z + U(Y_i \llcorner M)] = \mathsf{C}[M, \beta(M \lrcorner Y_i) Z]\,. \tag{8.55}$$

$\beta(M \lrcorner Y_i) M$ is the part of M which contributes to $\mathsf{C}(M, Y_i)$. This yields the decomposition:

$$\begin{aligned} \mathbb{V}(Y_i) &= \mathbb{V}[\beta(M \lrcorner Y_i) Z] + \mathbb{V}[U(Y_i \llcorner M)] \\ &= \beta(M \lrcorner Y_i)^2 \mathbb{V}(Z) + \mathbb{V}[U(Y_i \llcorner M)] \\ &= \frac{\mathsf{C}(M, Y_i)^2}{\mathbb{V}(M)} + \mathbb{V}[U(Y_i \llcorner M)] \\ &= \rho(M, Y_i)^2 \mathbb{V}(Y_i) + \mathbb{V}[U(Y_i \llcorner M)] \end{aligned} \tag{8.56}$$

where

$$\rho(M, Y_i) := \frac{\mathsf{C}(M, Y_i)}{\sigma(M)\sigma(Y_i)}\,, \tag{8.57}$$

$$\beta(M \lrcorner Y_i) = \rho(M, Y_i) \frac{\sigma(Y_i)}{\sigma(M)}\,, \tag{8.58}$$

hence

$$\rho(M, Y_i)^2 = 1 - \frac{\mathbb{V}[U(Y_i \llcorner M)]}{\mathbb{V}(Y_i)}\,. \tag{8.59}$$

It is interesting to see what happens to $\beta(M \lrcorner Y_i)$ when $\mathbb{V}(Y_i)$ is large or small $(1 \le i \le n)$. By the Cauchy-Schwarz inequality,

$$\mathsf{C}(M, Y_i)^2 \le \mathbb{V}(M)\mathbb{V}(Y_i)\,, \tag{8.60}$$

$$|\beta(M \lrcorner Y_i)| = \frac{|\mathsf{C}(M, Y_i)|}{\mathbb{V}(M)} = |\rho(M, Y_i)| \frac{\sigma(Y_i)}{\sigma(M)} \le \frac{\sigma(Y_i)}{\sigma(M)} \tag{8.61}$$

and

$$\begin{array}{ll} 0<\beta(M \dashv Y_i) \leq \frac{\sigma(Y_i)}{\sigma(M)} & \text{if } \mathsf{C}(M, Y_i)>0 \\ \beta(M \dashv Y_i)=0 & \text{if } \mathsf{C}(M, Y_i)=0 \\ -\frac{\sigma(Y_i)}{\sigma(M)} \leq \beta(M \dashv Y_i)<0 & \text{if } \mathsf{C}(M, Y_i)<0 . \end{array} \tag{8.62}$$

Note we can have $\sigma(Y_i) > \sigma(M)$, so $\beta(M \dashv Y_i)$ can be arbitrarily small (or large). Since $|\rho(M, Y_i)| \leq 1$, we have the inequality:

$$\mathbb{V}(M) = \sum_{i=1}^{n} \mathsf{C}(M, Y_i) = \sum_{i=1}^{n} \rho(M, Y_i)\sigma(M)\sigma(Y_i) \tag{8.63}$$

$$\leq \sigma(M)\sum_{i=1}^{n} \sigma(Y_i) \tag{8.64}$$

hence

$$\sigma(M) = \sum_{i=1}^{n} \rho(M, Y_i)\sigma(Y_i) \leq \sum_{i=1}^{n} \sigma(Y_i) . \tag{8.65}$$

For any i, we have:

$$\mathsf{C}(M - Y_i, Y_i) = \mathsf{C}(M, Y_i) - \mathsf{C}(Y_i, Y_i) = \mathsf{C}(M, Y_i) - \mathbb{V}(Y_i) , \tag{8.66}$$

$$\begin{aligned} \mathsf{C}(M, M - Y_i) &= \mathsf{C}[\mathsf{Y}_i + (M - Y_i), M - Y_i] \\ &= \mathsf{C}[\mathsf{Y}_i, M - Y_i] + \mathsf{C}[M - Y_i, M - Y_i] = \mathsf{C}[\mathsf{Y}_i, M - Y_i] + \mathbb{V}(M - Y_i) , \end{aligned} \tag{8.67}$$

hence

$$\mathsf{C}(M - Y_i, Y_i) = 0 \Leftrightarrow \mathsf{C}(M, Y_i) = \mathbb{V}(Y_i) \Leftrightarrow \mathsf{C}(M, M - Y_i) = \mathbb{V}(M - Y_i) . \tag{8.68}$$

If

$$\mathsf{C}(M - Y_i, Y_i) = 0 \tag{8.69}$$

for some i, we have:

$$\mathsf{C}(M, Y_i) = \mathbb{V}(Y_i) \geq 0 , \tag{8.70}$$

$$\mathsf{C}(M, M - Y_i) = \mathbb{V}(M - Y_i) \geq 0 , \tag{8.71}$$

$$\mathbb{V}(M) = \mathbb{V}[Y_i + (M - Y_i)] = \mathbb{V}(Y_i) + \mathbb{V}(M - Y_i) , \tag{8.72}$$

$$0 \leq \beta(M \dashv Y_i) = \frac{\mathbb{V}(Y_i)}{\mathbb{V}(M)} \leq 1 , \quad 0 \leq \beta(M \dashv M - Y_i) = \frac{\mathbb{V}(M - Y_i)}{\mathbb{V}(M)} \leq 1 , \tag{8.73}$$

$$\beta(M \rightharpoonup Y_i) + \beta(M \rightharpoonup M - Y_i) = 1\,. \tag{8.74}$$

If $Y_1, \ldots, Y_n$ are uncorrelated, *i.e.*

$$\mathsf{C}(Y_i, Y_j) = 0 \text{ for } i \neq j, \tag{8.75}$$

then, for $i, \ldots, n$,

$$\mathsf{C}(M - Y_i, Y_i) = 0\,, \tag{8.76}$$

$$\mathsf{C}(M, Y_i) = \mathbb{V}(Y_i)\,, \tag{8.77}$$

$$0 \leq \beta(M \rightharpoonup Y_i) = \frac{\mathbb{V}(Y_i)}{\mathbb{V}(M)} \leq 1\,, \tag{8.78}$$

$$\rho(M, Y_i)^2 = \frac{\mathsf{C}(M, Y_i)^2}{\mathbb{V}(M)\mathbb{V}(Y_i)} = \frac{\mathbb{V}(Y_i)}{\mathbb{V}(M)} = \beta(M \rightharpoonup Y_i)\,, \tag{8.79}$$

$$\sum_{i=1}^{n} \mathbb{V}(Y_i) = \mathbb{V}(M)\,, \tag{8.80}$$

$$\sum_{i=1}^{n} \beta(M \rightharpoonup Y_i) = 1\,. \tag{8.81}$$

Further, if $\mathbb{V}(Z) > 0$ and $\mathbb{V}[U(Y_i \leftharpoonup M)] > 0$, for $i = 1, \ldots, n$, we have:

$$\beta(M \rightharpoonup Y_i) = \frac{\sigma[U(Y_i \leftharpoonup M)]}{\sigma(M)}\, t(M, Y_i)\,, \quad \text{for } i = 1, \ldots, n, \tag{8.82}$$

$$\sum_{i=1}^{n} \frac{\sigma[U(Y_i \leftharpoonup M)]}{\sigma(M)}\, t(M, Y_i) = 1\,, \tag{8.83}$$

$$\sum_{i=1}^{n} \sigma[U(Y_i \leftharpoonup M)]\, t(M, Y_i) = \sigma(M)\,. \tag{8.84}$$

8.1.3. Covariance-variance subdecompositions

In (8.5), suppose that

$$Y_i = F_i + V_i\,, \quad i = 1, \ldots, n\,. \tag{8.85}$$

Then

$$M = F + V \tag{8.86}$$

where

$$F := \sum_{i=1}^{n} F_i\,, \quad V := \sum_{i=1}^{n} V_i\,, \tag{8.87}$$

and the covariance and linear regression coefficients are correspondingly decomposed:

$$\mathsf{C}(Z, Y_i) = \mathsf{C}(Z, F_i + V_i) = \gamma_i \mathsf{C}(Z, F_i) + \mathsf{C}(Z, V_i)\,, \quad i = 1, \ldots, n, \tag{8.88}$$

$$\beta(Z \rightharpoondown Y_i) = \beta(Z \rightharpoondown F_i) + (Z, V_i)\,, \quad i = 1, \ldots, n, \tag{8.89}$$

$$\mathsf{C}(Z, M) = \sum_{i=1}^{n} \mathsf{C}(Z, F_i) + \sum_{i=1}^{n} \mathsf{C}(Z, V_i)\,, \tag{8.90}$$

$$\beta(Z \rightharpoondown M) := \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \frac{\mathsf{C}(Z, Y_i)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \beta(Z \rightharpoondown Y_i)\,, \tag{8.91}$$

$$\beta(Z \rightharpoondown M) Z = \sum_{i=1}^{n} \beta(Z \rightharpoondown F_i) Z + \sum_{i=1}^{n} \beta(Z \rightharpoondown V_i) Z\,. \tag{8.92}$$

If $Z = M$ and $\mathbb{V}(M) > 0$, we get variance subdecompositions:

$$\begin{aligned} \mathbb{V}(M) &= \mathsf{C}(M, F + V) \\ &= \mathsf{C}(M, F) + \mathsf{C}(M, V)\,, \end{aligned} \tag{8.93}$$

$$\mathsf{C}(M, F) = \sum_{i=1}^{n} \mathsf{C}(M, F_i)\,, \tag{8.94}$$

$$\mathsf{C}(M, V) = \sum_{i=1}^{n} \mathsf{C}(M, V_i)\,, \tag{8.95}$$

$$\beta(M \rightharpoondown F) + \beta(M \rightharpoondown V) = 1\,, \tag{8.96}$$

$$\beta(M \rightharpoondown F) = \frac{\mathsf{C}(M, F)}{\mathbb{V}(M)} = \sum_{i=1}^{n} \beta(M \rightharpoondown F_i)\,, \tag{8.97}$$

$$\beta(M \rightharpoondown V) = \frac{\mathsf{C}(M, V)}{\mathbb{V}(M)} = \sum_{i=1}^{n} \beta(M \rightharpoondown V_i)\,, \tag{8.98}$$

$$\beta(M \rightharpoondown F) + \beta(M \rightharpoondown V) = 1\,. \tag{8.99}$$

$\mathsf{C}(M, F_i)$ is the contribution of F_i to the variance $\mathbb{V}(M)$, and $\mathsf{C}(M, V_i)$ is the contribution of V_i to $\mathbb{V}(M)$

8.2. Linear combination of random variables

Consider the weighted average

$$M := \sum_{i=1}^{n} \lambda_i X_i = \lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_n \tag{8.100}$$

where $\lambda_1, \ldots, \lambda_n$ are real constants. This is equivalent to considering (8.5) with

$$Y_i = \lambda_i X_i\,, \quad i = 1, \ldots, n\,, \tag{8.101}$$

so that the results of Sections 8.1.1 and 8.1.2 apply. In particular, for $i = 1, \ldots, n$,

$$\mathsf{C}(Z, Y_i) = \lambda_i \mathsf{C}(Z, X_i)\,, \quad \mathsf{C}(M, Y_i) = \lambda_i \mathsf{C}(M, X_i), \quad \mathbb{V}(Y_i) = \lambda_i^2\, \mathbb{V}(X_i)\,, \tag{8.102}$$

$$\beta(Z \dashv Y_i) := \frac{\lambda_i\, \mathsf{C}(Z, X_i)}{\mathbb{V}(Z)} = \lambda_i \beta(Z \dashv X_i)\,, \tag{8.103}$$

$$\beta(M \dashv Y_i) := \frac{\lambda_i\, \mathsf{C}(M, X_i)}{\mathbb{V}(M)} = \frac{\lambda_i \mathsf{C}(M, X_i)}{\mathbb{V}(X_i)} \frac{\mathbb{V}(X_i)}{\mathbb{V}(M)} = \lambda_i \beta(Z \dashv X_i) \frac{\mathbb{V}(X_i)}{\mathbb{V}(M)}\,. \tag{8.104}$$

For any random variable Z, we have:

$$\mathsf{C}(Z, M) = \sum_{i=1}^{n} \lambda_i \mathsf{C}(Z, X_i)\,, \tag{8.105}$$

$$\beta(Z \dashv M) := \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \lambda_i \frac{\mathsf{C}(M, X_i)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \lambda_i \beta(Z \dashv X_i)\,. \tag{8.106}$$

When Z and M are interchanged, we see that:

$$\beta(M \dashv Z) = \sum_{i=1}^{n} \lambda_i \frac{\mathsf{C}(X_i, Z)}{\mathbb{V}(M)} = \frac{\mathbb{V}(Z)}{\mathbb{V}(M)} \sum_{i=1}^{n} \lambda_i \beta(Z \dashv X_i) \tag{8.107}$$

where

$$\beta(Z \dashv X_i) = \frac{\mathsf{C}(Z, X_i)}{\mathbb{V}(Z)} = \rho(X_i, M)\, \sigma(X_i) \sigma(M)\,, \quad i = 1, \ldots, n\,. \tag{8.108}$$

Note also that

$$\beta(Z \dashv M) Z = \sum_{i=1}^{n} \beta(Z \dashv Y_i) Z = \sum_{i=1}^{n} \lambda_i \mathsf{C}(M, X_i) Z\,,\,, \tag{8.109}$$

$$M = \beta(Z \lrcorner M) Z + U(M \llcorner Z) = \sum_{i=1}^{n} \lambda_i \mathsf{C}(M, X_i) Z + U(M \llcorner Z)\,, \tag{8.110}$$

with

$$\mathsf{C}[Z, U(M \llcorner Z)] = \mathsf{C}[\beta(Z \lrcorner M) Z, U(M \llcorner Z)] = 0\,, \tag{8.111}$$

$$X_i = \beta(Z \lrcorner X_i) Z + U(X_i \llcorner Z)\,, \tag{8.112}$$

with

$$\mathsf{C}[Z, U(X_i \llcorner Z)] = \mathsf{C}[\beta(Z \lrcorner X_i) Z, U(X_i \llcorner Z)] = 0\,, \quad i = 1, \ldots, n\,, \tag{8.113}$$

$$\mathsf{C}(Z, X_i) = \mathsf{C}[Z, \beta(Z \lrcorner Y_i) Z + U(Z \llcorner X_i)] = \mathsf{C}[Z, \beta(Z \lrcorner X_i) Z]\,, \tag{8.114}$$

$$\begin{aligned}
\mathbb{V}(M) &= \mathbb{V}[\beta(Z \lrcorner M) Z] + \mathbb{V}[U(M \llcorner Z)] \\
&= \sum_{i=1}^{n} \lambda_i \beta(X_i, Z) \mathbb{V}(Z) + \mathbb{V}[U(M \llcorner Z)] \\
&= \sum_{i=1}^{n} \lambda_i \mathsf{C}(X_i, Z) + \mathbb{V}[U(M \llcorner Z)] \\
&= \sum_{i=1}^{n} \lambda_i \mathsf{C}(X_i, Z) + \mathsf{C}[M, U(M \llcorner Z)] \\
&= \mathsf{C}(M, Z) + \mathsf{C}[M, U(M \llcorner Z)]\,.
\end{aligned} \tag{8.115}$$

If

$$Z = M\,, \tag{8.116}$$

we see that:

$$\begin{aligned}
\mathbb{V}(M) &= \mathsf{C}(M, Y) \\
&= \sum_{i=1}^{n} \lambda_i \mathsf{C}(M, X_i) \\
&= \sum_{i=1}^{n} \lambda_i \beta(M \lrcorner X_i) \mathbb{V}(M)\,,
\end{aligned} \tag{8.117}$$

$$\sum_{i=1}^{n} \lambda_i \frac{\mathsf{C}(M, X_i)}{\mathbb{V}(M)} = \sum_{i=1}^{n} \lambda_i \beta(M \lrcorner X_i) = 1\,, \tag{8.118}$$

where $\lambda_i \beta(X_i \dashv M)$ may be negative.

Further,

$$\begin{aligned} \mathbb{V}(M) &= \sigma(M)^2 = \sum \lambda_i \rho(X_i, M)\, \sigma(X_i) \sigma(M) \\ &\leq \sum_{i=1}^{n} |\lambda_i|\, \sigma(X_i) \sigma(M)\,, \end{aligned} \tag{8.119}$$

$$\sigma(M) = \sum_{i=1}^{n} \lambda_i \rho(X_i, M)\, \sigma(X_i) \leq \sum_{i=1}^{n} |\lambda_i|\, \sigma(X_i) \tag{8.120}$$

and, using the Cauchy-Schwarz inequality,

$$\begin{aligned} \sigma(M) &= \sum_{i=1}^{n} \lambda_i \rho(X_i, M)\, \sigma(X_i) \\ &\leq \Big[\sum_{i=1}^{n} \lambda_i^2 \sigma(X_i)^2\Big]^{1/2} \Big[\sum_{i=1}^{n} \rho(X_i, M)^2\Big]^{1/2} \\ &\leq \sqrt{n} \Big[\sum_{i=1}^{n} \lambda_i^2 \sigma(X_i)^2\Big]^{1/2} = \sqrt{n} \Big[\sum_{i=1}^{n} \lambda_i^2\, \mathbb{V}(X_i)\Big]^{1/2}, \end{aligned} \tag{8.121}$$

$$\frac{1}{n} \mathbb{V}(M) \leq \sum_{i=1}^{n} \lambda_i^2\, \mathbb{V}(X_i)\,. \tag{8.122}$$

If $\lambda_i = 1$, for $i = 1, \ldots, n$, we have:

$$M := \sum_{i=1}^{n} X_i\,, \tag{8.123}$$

$$\mathbb{V}(M) = \sum_{i=1}^{n} \mathsf{C}(X_i, M)\,, \tag{8.124}$$

and, if $\mathbb{V}(M) > 0$,

$$\beta(M \dashv M) = \sum_{i=1}^{n} \beta(X_i \dashv M) = 1\,. \tag{8.125}$$

(8.124) can be interpreted as a decomposition of the variance of M in terms of its components M, and (8.125) as a regression decomposition.

If $\lambda_i = 1/n$, $i = 1, \dots, n$, we have:

$$\frac{1}{n}\sum_{i=1}^{n} \beta(X_i \rightharpoondown M) = 1 \quad \text{if } \mathbb{V}(M) > 0\,, \tag{8.126}$$

and

$$\sigma(M) \leq \Big[\sum_{i=1}^{n} \sigma(X_i)^2\Big]^{1/2} \tag{8.127}$$

Further,

$$\Big[\sum_{i=1}^{n} \lambda_i^2 \sigma(X_i)^2\Big]^2 \leq \Big[\sum_{i=1}^{n} \lambda_i^4\Big]\Big[\sum_{i=1}^{n} \sigma(X_i)^4\Big] \tag{8.128}$$

hence

$$\Big[\sum_{i=1}^{n} \lambda_i^2 \sigma(X_i)^2\Big]^{1/2} \leq \Big[\sum_{i=1}^{n} \lambda_i^4\Big]^{1/4}\Big[\sum_{i=1}^{n} \sigma(X_i)^4\Big]^{1/4}. \tag{8.129}$$

If $\sum_{i=1}^{n} \lambda_i = 1$, the maximum value of $\sum_{i=1}^{n} \lambda_i^4$ is achieved by taking $\lambda_i = 1/n$, $i = 1, \dots, n$. Thus

$$\sum_{i=1}^{n} \lambda_i^4 \leq \frac{1}{n^3}\,, \quad i = 1, \dots, n, \tag{8.130}$$

$$\Big[\sum_{i=1}^{n} \lambda_i^2 \sigma(X_i)^2\Big]^{1/2} \leq \frac{1}{n^{3/4}}\Big[\sum_{i=1}^{n} \sigma(X_i)^4\Big]^{1/4}, \tag{8.131}$$

$$\sigma(M) \leq \frac{1}{n^{1/4}}\Big[\sum_{i=1}^{n} \sigma(X_i)^4\Big]^{1/4} = \Big[\frac{1}{n}\sum_{i=1}^{n} \sigma(X_i)^4\Big]^{1/4}, \tag{8.132}$$

$$\mathbb{V}(M) \leq \Big[\frac{1}{n}\sum_{i=1}^{n} \sigma(X_i)^4\Big]^{1/2} \tag{8.133}$$

8.3. Linear combination of random variables with disturbance

Suppose

$$Y = X_1 + X_2 + \cdots + X_n + U = M + U \tag{8.134}$$

where

$$M := \sum_{i=1}^{n} X_i \tag{8.135}$$

and $Y, X_1, \dots, X_n, U$ are random variables with finite second moments. Then, if Z is a also a random variable with finite second moment, we have:

$$\mathsf{C}(Z,Y) = \sum_{i=1}^{n} \mathsf{C}(Z, X_i) + \mathsf{C}(Z, U)\,, \tag{8.136}$$

$$\beta(Z{\lnot}Y) = \sum_{i=1}^{n} \beta(Z{\lnot}X_i) + \beta(Z{\lnot}U)\,. \tag{8.137}$$

The above equation provides a decomposition of the covariance $\mathsf{C}(Y, Z)$ and $\beta(Y{\lnot}Z)$. For $Z = Y$, we get:

$$\mathbb{V}(Y) = \sum_{i=1}^{n} \mathsf{C}(Y, X_i) + \mathsf{C}(Y, U) \tag{8.138}$$

which provides a decomposition of the variance $\mathbb{V}(Y)$, and

$$\sum_{i=1}^{n} \beta(Y{\lnot}X_i) + \beta(Y{\lnot}U) = 1\,. \tag{8.139}$$

$\beta(Y{\lnot}X_i)$ is then the proportional contribution of X_i to the variance of Y. If $\mathsf{C}(U, M) = 0$, we have:

$$\mathbb{V}(Y) = \sum_{i=1}^{n} \mathsf{C}(X_i, Y) + \mathbb{V}(U)\,, \tag{8.140}$$

$$\sum_{i=1}^{n} \beta(Y{\lnot}X_i) = 1 - \frac{\mathbb{V}(U)}{\mathbb{V}(Y)}\,. \tag{8.141}$$

If furthermore $U = 0$, we have:

$$\mathbb{V}(Y) = \sum_{i=1}^{n} \mathsf{C}(X_i, Y)\,, \tag{8.142}$$

$$\sum_{i=1}^{n} \beta(Y{\lnot}X_i) = 1\,. \tag{8.143}$$

8.4. Factor decompositions

In (8.5), suppose that

$$Y_i = \gamma_i F_i + V_i\,, \tag{8.144}$$

$$\mathsf{C}(F_i, V_i) = 0\,, \tag{8.145}$$

for some $i \in \{1, \ldots, n\}$., where

For any random variable Z, we have:

$$\mathsf{C}(Z, Y_i) = \mathsf{C}(Z, \gamma_i F_i + V_i) = \gamma_i \mathsf{C}(Z, F_i) + \mathsf{C}(Z, V_i)\,, \tag{8.146}$$

$$\beta(Z \dashv Y_i) = \gamma_i \beta(Z \dashv F_i) + \beta(Z \dashv V_i)\,. \tag{8.147}$$

In particular, for $Z = M$,

$$\mathsf{C}(M, Y_i) = \mathsf{C}(M, \gamma_i F_i + V_i) = \gamma_i \mathsf{C}(M, F_i) + \mathsf{C}(M, V_i)\,, \tag{8.148}$$

$$\beta(M \dashv Y_i) = \gamma_i \beta(M \dashv F_i) + \beta(M \dashv V_i)\,. \tag{8.149}$$

If furthermore

$$\mathsf{C}(M - Y_i, F_i) = \mathsf{C}(M - Y_i, V_i) = 0\,, \tag{8.150}$$

we have:

$$\begin{aligned}
\mathsf{C}(M, Y_i) &= \mathsf{C}[Y_i + (M - Y_i), \gamma_i F_i] + \mathsf{C}[Y_i + (M - Y_i), V_i] \\
&= \gamma_i \mathsf{C}[Y_i, F_i] + \mathsf{C}[Y_i, V_i] \\
&= \gamma_i \mathsf{C}[\gamma_i F_i + V_i\,, F_i] + \mathsf{C}[\gamma_i F_i + V_i\,, V_i] \\
&= \gamma_i^2 \mathbb{V}(F_i) + \mathbb{V}(V_i) \geq 0\,,
\end{aligned} \tag{8.151}$$

$$\begin{aligned}
\beta(M \dashv Y_i) &= \beta(M \dashv F_i) + \beta(M \dashv V_i) \\
&= \gamma_i^2 \frac{\mathbb{V}(F_i)}{\mathbb{V}(M)} + \frac{\mathbb{V}(V_i)}{\mathbb{V}(M)} \geq 0\,,
\end{aligned} \tag{8.152}$$

$$\begin{aligned}
\mathbb{V}(M) &= \sum_{i=1}^{n} \mathsf{C}(M, Y_i) \\
&= \sum_{i=1}^{n} [\gamma_i^2 \mathbb{V}(F_i) + \mathbb{V}(V_i)]
\end{aligned} \tag{8.153}$$

$\gamma_i^2 \mathbb{V}(F_i)$ represents the contribution of F_i to $\mathbb{V}(M)$, that goes through Y_i.

Suppose

$$Y_i = \gamma_i F + V_i, \quad i = 1, \ldots, n\,, \tag{8.154}$$

where $\gamma_1, \ldots, \gamma_n$ are real constants and

$$\mathsf{C}(F, V_i) = 0, \quad i = 1, \ldots, n, \tag{8.155}$$

$$\mathsf{C}(V_i, V_j) = 0, \quad \text{for } i \neq j. \tag{8.156}$$

Then

$$\sum_{i=1}^{n} Y_i = \bar{\gamma} F + V \sum_{i=1}^{n} V_i \tag{8.157}$$

where

$$\bar{\gamma} := \sum_{i=1}^{n} \gamma_i, \quad V := \sum_{i=1}^{n} V_i. \tag{8.158}$$

For any random variable Z, we have:

$$\mathsf{C}(Z, Y_i) = \mathsf{C}(Z, \gamma_i F + V_i) = \gamma_i \mathsf{C}(Z, F) + \mathsf{C}(Z, V_i), \quad i = 1, \ldots, n, \tag{8.159}$$

$$\begin{aligned} \mathsf{C}(Z, M) &= \mathsf{C}\Big(Z, \sum_{i=1}^{n} Y_i\Big) = \sum_{i=1}^{n} \mathsf{C}(Z, Y_i) \\ &= \bar{\gamma} \mathsf{C}(Z, F) + \sum_{i=1}^{n} \mathsf{C}(Z, V_i), \end{aligned} \tag{8.160}$$

$$\begin{aligned} \beta(Z \rightharpoonup M) &= \frac{\mathsf{C}(Z, M)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \frac{\mathsf{C}(Z, Y_i)}{\mathbb{V}(Z)} = \sum_{i=1}^{n} \beta(Z \rightharpoonup Y_i) \\ &= \bar{\gamma} \frac{\mathsf{C}(Z, F)}{\mathbb{V}(Z)} + \sum_{i=1}^{n} \frac{\mathsf{C}(Z, V_i)}{\mathbb{V}(Z)} \\ &= \bar{\gamma} \beta(Z \rightharpoonup F) + \sum_{i=1}^{n} \beta(Z \rightharpoonup V_i). \end{aligned} \tag{8.161}$$

Equation (8.159) shows that $\mathsf{C}(Z, Y_i)$ has two components: one associated with the common factor F, and another one associated with the idiosyncratic factor V_i.

9. Sources and additional references

Good overviews of various notions associated with covariances, correlations and regression may be found in Hannan (1970, Chapter 1), Theil (1971, Chapter 4), Kendall and Stuart (1979, Chapters 26-28), Rao (1973, Section 4g), Drouet Mari and Kotz (2001), and Anderson (2003, Chapter 1). See also Lehmann (1966).

References

ANDERSON, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, third edn.

DROUET MARI, D., AND S. KOTZ (2001): *Correlation and Dependence*. World Scientific, River Edge, New Jersey.

HANNAN, E. J. (1970): *Multiple Time Series*. John Wiley & Sons, New York.

KENDALL, M., AND A. STUART (1979): *The Advanced Theory of Statistics. Volume 2: Inference and Relationship*. Macmillan, New York, fourth edn.

LEHMANN, E. L. (1966): “Some Concepts of Dependence,” *The Annals of Mathematical Statistics*, 37, 1137–1153.

RAO, C. R. (1973): *Linear Statistical Inference and its Applications*. John Wiley & Sons, New York, second edn.

THEIL, H. (1971): *Principles of Econometrics*. John Wiley & Sons, New York.