ERASMUS B4 Course

Winter Semester 2020/2021, January 7

Subject: Introduction to the Covariance Calculus, part I

Suppose we have a R.V.

$$\Omega \ni \omega \longrightarrow (X,Y)(\omega) \in \mathbb{R}^2$$

Def 1 (the covariance of (X,Y))

By the covariance of X & Y we understand the number

$$\operatorname{cov}(X,Y) \stackrel{df}{=} E\big((X-EX)(Y-EY)\big)$$

Since

$$E\big((X-EX)(Y-EY)\big) = E\big(XY - XEY - YEX + (EX)(EY)\big)$$

$$= E(XY) - (EX)(EY) - (EY)(EX) + (EX)(EY),$$

we see that

#1 $$\underline{\operatorname{cov}(X,Y) = E(XY) - (EX)(EY)},$$

and therefore we have

1°. If $Y = X$, then

$$\operatorname{cov}(X,X) = \operatorname{var}(X)$$

2°. If X & Y are stochastically independent, then $E(XY) = (EX)(EY)$, and consequently $\mathrm{cov}(X,Y) = 0$, but as we see later not vice-versa!

3°. From general theory of Lebesgue integral we have

$$E(XY) \leq \sqrt{E(X^2)}\left(\sqrt{E(Y^2)}\right),$$

so if X and Y have 2^{nd} moment, then

~~[illegible]~~ $\mathrm{cov}(X,Y)$ exists.

Example 1.

Suppose that for (X,Y) we have

$$(X,Y) \longrightarrow \mathbb{P} \in \mathbb{M}_{n \times m}, \text{ where}$$

$$\mathbb{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 1/4 & 0 \\ 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 \end{bmatrix} \end{array}\begin{array}{c} \\ 1/4 \\ 1/2 \\ 1/4 \end{array} : d(X)$$

$d(Y)$: $1/4 \quad 1/2 \quad 1/4$,

so $X(\Omega) = Y(\Omega) = \{0,1,2\}$, and $d(X) = d(Y)$.

Therefore X & Y are not stochast. independent (WHY?)

We will calculate the $\mathrm{cov}(X,Y)$:

We have:

a) $EX = 0 \cdot 1/4 + 1 \cdot 1/2 + 2 \cdot 1/4 = 1 = EY$

b) since $(XY)(\Omega) = \{0,1,2,4\}$, from the joint distribution of (X,Y)

$$d(XY): \begin{array}{c|c|c|c} 0 & 1 & 2 & 4 \\ \hline 1/2 & 0 & 1/2 & 0 \end{array} = \begin{array}{c|c} 0 & 2 \\ \hline 1/2 & 1/2 \end{array},$$

and $E(XY) = 0 \cdot 1/2 + 2 \cdot 1/2 = 1$.

Finally, $\mathrm{cov}(X,Y) = 1 - 1 = 0$

4°. We know, that if X & Y are s.i. then

$$var(X+Y) = var(X) + var(Y).$$

Now, suppose that X & Y are arbitrary. Then

$$var(X+Y) = E((X+Y)^2) - (E(X+Y))^2 =$$

$$= E(X^2+2XY+Y^2) - ((EX)^2 + 2(EX)(EY) + (EY)^2).$$

Therefore in general case

(#2) $$var(X+Y) = (E(X^2) - (EX)^2) + (E(Y^2) - (EY)^2) + 2(E(XY) - (EX)(EY)),$$

so from (#2)

(#3) $$\boxed{var(X+Y) = var(X) + var(Y) + 2cov(X,Y)},$$

Later, we will see that $cov(X,Y)$ is a <u>kind of</u> <u>measure</u> <u>of dependency</u> of X & Y

Example 2.

Suppose we observe some population of people consists of 5.000. As the result of our observation is hair color and eyes color of observed people.

The hair colour is: blond and dark

eyes colour is: blue, green and gray.

The results of these observations are as follows:

Eye colour	Hair color	
	blond	dark
blue	1100	500
green	800	1300
gray	300	1000

We will give the probabilistic description of the above observation.

If we use the notion of frequence of observated phenomenos: $\frac{n}{N}$, when N - the amount of the whole group, n - the amount of the given subgroup, we obtain the following matrix

$$\mathbb{P} = \begin{bmatrix} \frac{11}{50} & \frac{1}{10} \\ \frac{8}{50} & \frac{13}{50} \\ \frac{3}{50} & \frac{1}{5} \end{bmatrix}.$$

From general theory of joint dist. it means that on some K.P.M (Ω, Σ, P) we have two r.v. X, Y, and R.V. (X, Y),

where X means the eyes color,

Y means the hair color,

$$(X, Y) \longrightarrow \mathbb{P},$$

$$d(X) = \left(\frac{16}{50}, \frac{21}{50}, \frac{13}{50}\right)$$

$$d(Y) = \left(\frac{22}{50}, \frac{28}{50}\right)$$

Since

$$P(\{\omega \in \Omega : X(\omega) = \text{„blue"} \;\&\; Y(\omega) = \text{„blond"}\}) = \frac{11}{50}$$

and

$$P(\{\omega \in \Omega : X(\omega) = \text{„blue"}\}) \cdot P(\{\omega \in \Omega : Y(\omega) = \text{„blond"}\})$$

$$= \frac{16}{50} \cdot \frac{22}{50} = \frac{88}{25^2}, \text{ we see that}$$

X & Y are not independent

TASK 1.

Calculate $cov(X, Y)$.

Def 3.

If $cov(X, Y) > 0$ we say that r.v. X & Y are positively correlated,

$cov(X, Y) < 0$, negatively correlated.

In the case than $cov(X, Y)$ we say that X & Y are uncorrelated.

So, if X & Y are s.i. than are uncorrelated but not vice-versa!

Later we will improve the definition of $cov(X, Y)$ as a measure of dependency.

the concept of the correlation coefficient.

Assumptions.

X & Y has a second moment and

$var(X)\, var(Y) \neq 0$, so X and Y are not constant.

Def 2 (correlation coefficient)

$$\varrho \stackrel{df}{=} \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} \quad \left(= \varrho(X,Y) \right)$$

Remarks

1°. $\varrho(X,Y) = cov\left(\frac{X}{\sqrt{var X}}, \frac{Y}{\sqrt{var Y}} \right)$

2°. $\varrho(X,X) = \frac{var(X)}{var(X)} = 1$.

3°. Suppose that for (X,Y) we have

$P(\{\omega \in \Omega : Y(\omega) = aX(\omega) + b\}) = 1$, $a \neq 0$, $b \in \mathbb{R}$

and X has a second moment.

Then the $cov(X,Y)$ exists, and from (#1)

$$cov(X,Y) = E(XY) - (EX)(EY) =$$

$$= E(X(aX+b)) - (EX)(E(aX+b))$$

$$= aE(X^2) + bEX - a(EX)^2 - bEX,$$ so

$$cov(X,Y) = a(E(X^2) - (EX)^2) + b(EX - EX)$$

$$= a\,var(X).$$

But $var(Y) = a^2 var(X)$, so finally

$$\rho = \frac{cov(X,Y)}{\sqrt{var X}\sqrt{var Y}} = \frac{a\,var(X)}{|a|\,var(X)} = \frac{a}{|a|}$$

and we can write

$$\rho_{(X,aX+b)} = \begin{cases} 1, & a > 0 \\ -1, & a < 0 \end{cases}$$

Now we will <u>establish the spectrum of ϱ</u>.

First we note that for $a, b \in \mathbb{R}_+$

$$\varrho(aX, bY) = \frac{E(abXY) - E(aX)E(bY)}{\sqrt{var(aX)}\ \sqrt{var(bY)}} =$$

$$= \frac{ab\left(E(XY) - (EX)(EY)\right)}{ab\sqrt{var X}\cdot\sqrt{var Y}} = \varrho(X, Y)$$

In particular, if we put $a = \frac{1}{\sqrt{var X}}$, $b = \frac{1}{\sqrt{var Y}}$ we get

$$\varrho\left(\frac{X}{\sqrt{var X}}, \frac{Y}{\sqrt{var Y}}\right) = \varrho(X, Y) \left(= cov\left(\frac{X}{\sqrt{var X}}, \frac{Y}{\sqrt{var Y}}\right)\right)$$

On the other hand, from the formula (#)

$$0 \leq var\left(\frac{X}{\sqrt{var X}} + \frac{Y}{\sqrt{var Y}}\right) =$$

$$= var\left(\frac{X}{\sqrt{var X}}\right) + var\left(\frac{Y}{\sqrt{var Y}}\right) + 2cov\left(\frac{X}{\sqrt{var X}}, \frac{Y}{\sqrt{var Y}}\right)$$

$$\doteq 1 + 1 + 2\varrho(X, Y).$$

Therefore

$$\varrho(X,Y) \geq -1 \text{ for every } X \,\&\, Y$$

(with second moment!)

Now, if we change X onto $-X$, we obtain

$$\varrho(-X,Y) \geq -1$$
$$\varrho(-X,Y) = -\varrho(X,Y),$$

and $\varrho(X,Y) \leq 1$

Now, we are ready to formulate the main <u>theorem</u> on ϱ.

Theorem.

Let X & Y have at least second moment and nonconstant.

then:

1°. For $\rho_0 \overset{df}{=} |\rho| = |\rho(X,Y)|$

$$\rho_0 \in [0,1]$$

2°. If X & Y are s.ind. then

$$\rho_0 = 0.$$

So, if $\rho_0 > 0 \Rightarrow X$ & Y are not s.ind.

and ~~we can say that~~ for this reason

we can say that ρ_0 is the measure of dependency of X & Y.

3°. If $Y = aX + b$ $(a \neq 0)$ with prob. 1, then $\rho_0 = 1$ (it is our previous result!)

and vice-versa!

In such a situation we say that we have a linearity connection between X and Y

Since $\max \rho_0 = 1$, the linearity connection is the stronger version of stochastic dependency.

Example 3.

For $X(\Omega) = Y(\Omega) = \{0, 1, 2\}$ let

$$(X, Y) \longrightarrow \Pi, \quad \text{where}$$

$$\Pi = \begin{bmatrix} 0 & 1/4 & 0 \\ 1/4 & 0 & 1/4 \\ 0 & 1/4 & 0 \end{bmatrix}$$

We check if for X and Y ~~are~~ we have a linearity connection.

To do this, we use the above theorem.

Since $d_X : (1/4, 1/2, 1/4)$

$d_Y : (1/4, 1/2, 1/4)$

X and Y are not S.Ind.

Now $EX = EY = 1$ and

$XY(\Omega) = \{0, 1, 2, 4\}$, so

$$d_{XY} \begin{array}{c|c} 0 & 2 \\ \hline 1/2 & 1/2 \end{array}, \text{ so } E(XY) = 1.$$

Hence $cov(X,Y) = 0 \Rightarrow \rho_0 = 0$.

So we have an opposite to linearly connection property between X & Y.

TASK 2

Calculate ρ_0 for date given in example 2

Remark on covariance matrix

For given two variables X, Y with (at least second moment), we can define the matrix R as follows:

$$R = \begin{bmatrix} cov(X,X), & cov(X,Y) \\ cov(Y,X), & cov(Y,Y) \end{bmatrix}$$

Remark

In fact, in above we have a R.V. (X,Y). In trivially case, so if we have only a single r.v.

$$R = [\, cov(X,X) \,] = [\, var\, X \,].$$

Therefore, the correspondence

$(X,Y) \longrightarrow R$ is a generalization of variance for R.V.

Theorem (on R).

For every matrix R we have:

1) R is symmetric, so

$$R = R^T \text{ (transposition)}$$

2) $\det R \geq 0$

3) $\det R = 0$ iff X & Y are constant.

TASK 3

Prove this theorem!