Discussiones Mathematicae
Probability and Statistics 35 (2015) 57–60
doi:10.7151/dmps.1177

A NOTE ON CORRELATION COEFFICIENT BETWEEN RANDOM EVENTS

Czesław Stępniak

Department of Differential Equations and Statistics
Faculty of Mathematics and Natural Sciences
University of Rzeszów
Pigonia 1, 35–959 Rzeszów, Poland

e-mail: stepniak@umcs.lublin.pl

Abstract

Correlation coefficient is a well known measure of (linear) dependence between random variables. In his textbook published in 1980 L.T. Kubik introduced an analogue of such measure for random events A and B and studied its basic properties. We reveal that this measure reduces to the usual correlation coefficient between the indicator functions of A and B. In consequence the resuts by Kubik are obtained and strenghted directly. This is essential because the textbook is recommended by many universities in Poland.

Keywords: correlation coefficient between random events, correlation coefficient for random variables, synergy phenomenon.

2010 Mathematics Subject Classification: Primary: 62H20, Secondary: 62J05.

1. Correlation coefficient between random events

Correlation coefficient, called also Pearson's coefficient, is a well known measure of (linear) dependence between random variables X and Y. It may be defined as

$$\rho_{X,Y} = \frac{E[(X - EX)(Y - EY)]}{\sqrt{E(X - EX)^2 E(Y - EY)^2}}$$

providing the denominator does not vanish.

In his textbook ([2], p. 128–129) published in 1980 L.T. Kubik introduced an analogous measure for random events and studied its properties. After a slightly intricate argumentation he defined so called *correlation coefficient between the random events* A and B as

$$\rho(A,B) = \frac{P(A\cap B) - P(A)P(B)}{\sqrt{P(A)[1-P(A)]P(B)[1-P(B)]}}, \quad \text{if } P(A)[1-P(A)]P(B)[1-P(B)] \neq 0. \tag{1}$$

Kubik proved that this coefficient possesses the following properties:

1° $\rho(A,B) = \rho(B,A)$,

2° $-1 \leq \rho(A,B) \leq 1$,

3° $\rho(A,B) = 0$ iff A and B are independent,

4° If $\rho(A,B) = 1$ then $P(A\cap B) = P(A) = P(B)$,

5° If $\rho(A,B) = -1$ then $P(A\cap B) = 0$.

It is worth to note that if $P(A)[1-P(A)]P(B)[1-P(B)] = 0$ then the random events A and B are independent, while (1) is not defined. Therefore it would be more safely to complete the formula (1) by

$$\rho(A,B) = 0, \quad \text{if } P(A)[1-P(A)]P(B)[1-P(B)] = 0. \tag{2}$$

Then the property 3° holds.

First we reveal that $\rho(A,B)$ coincides with the usual correlation coefficient $\rho_{X,Y}$ between the random variables

$$X(\omega) = \mathbf{1}_A(\omega)$$

and

$$Y(\omega) = \mathbf{1}_B(\omega),$$

where symbol $\mathbf{1}_A$ stands for the indicator function of the set A, i.e.,

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

Thus the results $1° - 5°$ may be obtained directly from the well known properties:

(1) $\rho_{X,Y} = \rho_{Y,X}$,

(2) $-1 \leq \rho_{X,Y} \leq 1$,

(3) If X and Y are independent then $\rho_{X,Y} = 0$,

(4) If $|\rho_{X,Y}| = 1$ then $P[a(X - EX) = b(Y - EY)] = 1$ for some scalars a and b not both null,

(5) $sign(\rho_{X,Y}) = sign(ab)$, with a and b appearing in (4).

(see, for instance, ([1], p. 101), or ([4], p. 133)).

This information is essential because the textbook [2] is recommended by many universities in Poland.

In order to show that $\rho(A, B) = \rho_{X,Y}$ we only need to note that

$$\begin{aligned} EX &= EX^2 = P(A), \\ EY &= EY^2 = P(B), \end{aligned}$$

and

$$EXY = P(A \cap B).$$

In consequence,

$$\begin{aligned} E(X - EX)^2 &= EX^2 - (EX)^2 = P(A) - [P(A)]^2 = P(A)[1 - P(A)], \\ E(Y - EY)^2 &= EY^2 - (EY)^2 = P(B) - [P(B)]^2 = P(B)[1 - P(B)] \end{aligned}$$

and

$$E[(X - EX)(Y - EY)] = EXY - EXEY = P(A \cap B) - P(A)P(B).$$

It appears that in our convention (2) the converse statement to 4° is also true. In this situation the both statements 4° − 5° may be strengthened as below.

Lemma 1. $\rho(A, B) = 1$ *iff arbitrary of the following conditions holds:*

(a) $0 < P(A) < 1$ *and* $P[(A \setminus B) \cup (B \setminus A)] = 0$,

(b) $0 < P(A) < 1$ *and* $P(A \cap B) = P(A) = P(B) < 1$.

Proof. Equivalence of (a) and (b) is evident. Necessity of (b) is stated in 4° while its sufficiency may be verified directly. ■

Lemma 2. $\rho(A, B) = -1$ *iff arbitrary of the following conditions holds:*

(c) $0 < P(A) < 1$ *and* $P[(A \setminus B) \cup (B \setminus A)] = 1$,

(d) $0 < P(A) < 1$, $P(A \cap B) = 0$ *and* $P(A \cup B) = 1$,

(e) $P(B) = 1 - P(A) \neq 0$ *or* 1.

Proof. Equivalence of the conditions (c), (d) and (e) and sufficiency of (d) is evident. Necessity of $P(A \cap B) = 0$ is stated in 4°. For the necessity of $P(A \cup B) = 1$, suppose, by contradiction, that $P(B) < 1 - P(A)$. Then

$$\rho(A, B) = -\sqrt{\frac{P(A)}{1 - P(B)}}\sqrt{\frac{P(B)}{1 - P(A)}} > -1$$

This completes the proof of the lemma. ■

At the end let us mention about an interesting relation between the *sample* correlation coefficients and so called synergy phenomenon in regression model $\mathbf{y} = \mu\mathbf{1} + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \mathbf{e}$ with the response vector $\mathbf{y}$ and two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, of the explanatory variables. The synergy problem refers to a rather unexpected situation when the determination coefficient R^2 is greater than the sum $r^2_{x_1,y}$ and $r^2_{x_2,y}$ of the squares of the sample correlation coefficients between the response vector $\mathbf{y}$ and each of the explanatory vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. It was proved in [3] that a model is lack of synergy for all possible $\mathbf{y}$ if and only if $r^2_{x_1,x_2} = 0$ or 1.

Acknowledgement

This work was partially supported by the Centre for Innovation and Transfer of Natural Sciences and Engineering Knowledge.

References

[1] P. Brémaud, An Introduction to Probabilistic Modeling (Springer-Verlag, New York, 1994).

[2] L.T. Kubik, Probability. A Textbook for Teaching Mathematics Studies (in Polish) (Polish Scientific Publishers, Warsaw, 1980).

[3] K. Maciąg and C. Stępniak, *Lack of synergy regression models*, Comm. Statist. Theory Methods **43** (2014) 383–391.

[4] D. Stirzaker, Elementary Probability (Cambridge Univ. Press, Cambridge, 1994).

Received 20 March 2015