Difference between the notion of causation and Pearson correlation in a multivariate Gaussian context

Ait-taleb Nabil *

Abstract

In a Gaussian multivariate context, we will describe the steps to follow to differentiate the notion of Pearson correlation and the causality. This paper includes numerical examples clearly showing the difference between the two notions.

*Corresponding author: nabiltravail1982@gmail.com

1 Introduction

In this paper, we will understand in a Gaussian context from a proof how to relate the notion of causation to square of the multiple correlation.

For this, I will have to introduce the causal effect vector $\Delta E = E[X|\Omega] - E[X]$.

$X - \Delta E$ will correspond to the signal obtained when the causes Ω acts on X.

The unimpacted signal X and the impacted signal $X - \Delta E$ will be related to square of the multiple correlation $K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X}$.

To facilitate our understanding of correlation and causation, I will present a table that showcases the magnitude of correlations alongside their corresponding causation levels. Using the Seatblet matrix example found in the R software, we will expose the signals unimpacted and impacted by the causes.

Subsequently, we explore a scenario involving two causes acting on a variable. By delineating the correlation pairs associated with strong and weak causation, we shed light on the intricate relationship between these factors.

The paper concludes with numerical applications, specifically addressing a problem where two causes influence a variable. Through these examples, we establish connections between strong and weak correlations and the likelihood of causation.

2 Square of multiple correlation and causation

The relationship which links the causality to the correlations can be written as follows:

$$\boxed{K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} = 1 - \frac{Var(X - \Delta E)}{Var(X)}}$$

where $\#\Omega \geq 2$, $E(.|.)$ is the Gaussian conditional average and $Var(.)$ is the variance.

X is the unimpacted signal.

$\Delta E = E[X|\Omega] - E[X]$ is the causal effect vector.

$X - \Delta E$ corresponds to impacted signal obtained when the causes Ω act on X.

$0 \leq \frac{Var(X-\Delta E)}{Var(X)} \leq 1$ is the causal effect ratio of causes Ω acting on the variable X.

$0 \leq K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} \leq 1$ corresponds to square multiple correlation.

Proof:

In what follows, we will factorize the variance Σ_{X^2} of the conditional variance $\Sigma_{X^2|\Omega}$ to show the correlations K:

$\Sigma_{X^2|\Omega} = \Sigma_{X^2} - \Sigma_{X,\Omega}.\Sigma_{\Omega^2}^{-1}.\Sigma_{X,\Omega}$

$\Sigma_{X^2|\Omega} = \Sigma_{X^2} - \Sigma_{X,\Omega}.(diag^{-1}(\Sigma_{\Omega^2}))^{\frac{1}{2}}.K_{\Omega^2}^{-1}.(diag^{-1}(\Sigma_{\Omega^2}))^{\frac{1}{2}}.\Sigma_{\Omega,X}$

$\Sigma_{X^2|\Omega} = \Sigma_{X^2} - \Sigma_{X^2}^{\frac{1}{2}}.K_{X,\Omega}.K_{\Omega^2}^{-1}.\Sigma_{X^2}^{\frac{1}{2}}.K_{\Omega,X}$

$\Sigma_{X^2|\Omega} = \Sigma_{X^2}.(1 - K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X})$

The relationship can also be written:

$$K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} = 1 - \frac{\Sigma_{X^2|\Omega}}{\Sigma_{X^2}} = 1 - \frac{||X - E(X|\Omega)||^2}{||X - E(X)||^2} = 1 - \frac{\frac{||X-E(X|\Omega)||^2}{N}}{\frac{||X-E(X)||^2}{N}}$$

As we have:$E_\Omega(E(X|\Omega)) = \frac{1}{N}\sum_{\Omega} E(X|\Omega) = E(X)$, we obtain:

$K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} = 1 - \frac{Var(X-E[X|\Omega])}{Var(X)}$

By adding the constant $E[X]$ into the variance $Var(X - E[X|\Omega])$: $Var(X - E[X|\Omega]) = Var(X - (E[X|\Omega] - E[X]))$. By putting $\Delta E = E[X|\Omega] - E[X]$, we obtain the relationship: $K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} = 1 - \frac{Var(X-\Delta E)}{Var(X)}$

Note that the Gaussian entropy of X gives $h(X)$ and that the Gaussian entropy of the impacted signal $X - \Delta E$ gives the following Gaussian conditional entropy $h(X|\Omega)$:

$h(X - \Delta E) = \frac{1}{2}\ln(2.\pi.e.Var(X - (E[X|\Omega] - E[X]))) = \frac{1}{2}\ln(2.\pi.e.\Sigma_{X^2|\Omega}) = h(X|\Omega)$.

The signal X having an average $E[X]$ and an entropy $h(X)$ impacted by the causes Ω becomes the signal $X - \Delta E$ having an average $E[X]$ and a conditional entropy $h(X|\Omega)$.

3 Pearson correlation value range

We will explain the importance of correlations to interpret the order of magnitude in what will follow:

Level of correlation	ρ_{min}	ρ_{max}
Strong positive correlation	0.6	1
Moderate positive correlation	0.4	0.59
Weak positive correlation	0.2	0.39
Very Weak positive correlation	0	0.19
Strong negative correlation	-1	-0.6
Moderate negative correlation	-0.59	-0.4
Weak negative correlation	-0.39	-0.2
Very Weak negative correlation	-0.19	0

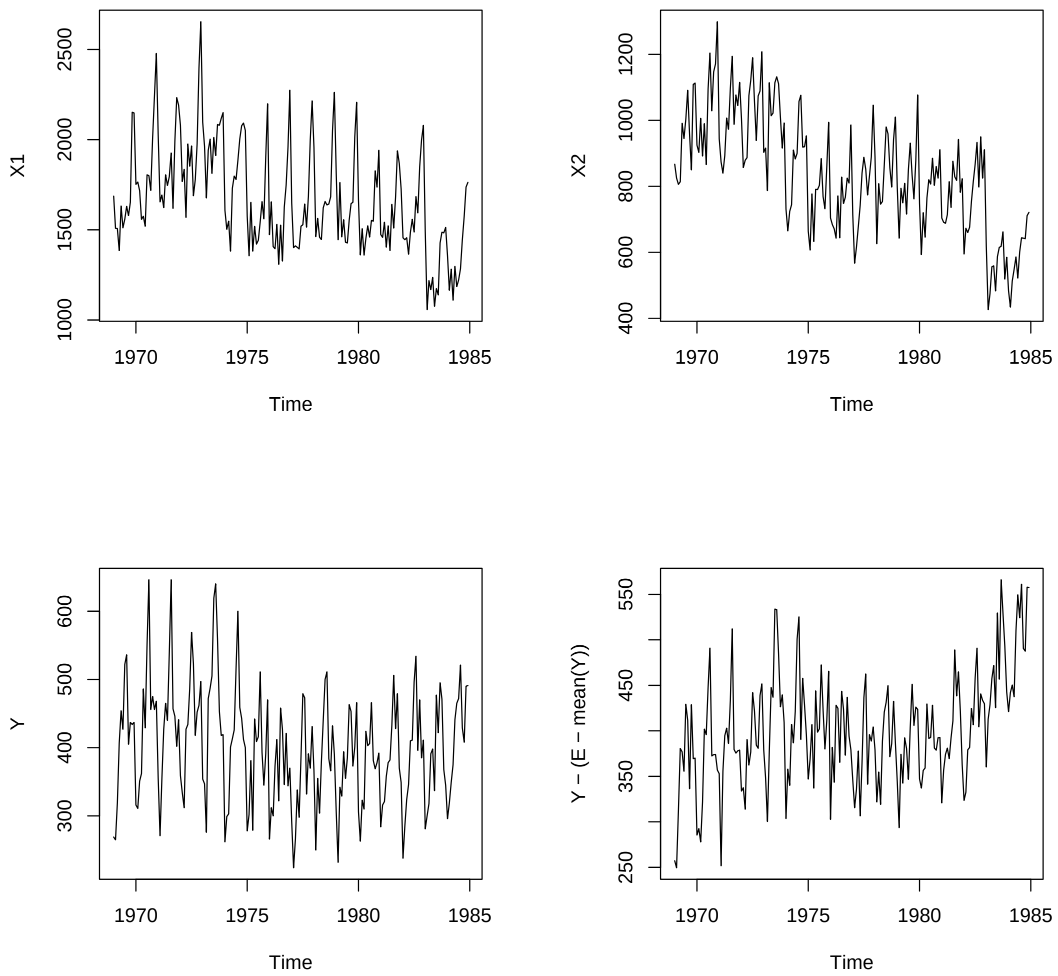
4 Causation value range

We will present a table containing the magnitudes of causation:

Level of causation	$min(K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X})$	$max(K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X})$
Very weak	0	0.199
Weak	0.2	0.399
Medium	0.4	0.599
Strong	0.6	0.799
Very Strong	0.8	1

5 Impact for Seatbelt Matrix Signals

In what follows, we will consider the Seatbelt matrix found in the R software to show the cause signals X_1, X_2 ,the signal unimpacted Y and the signal $Y - \Delta E$ impacted by the causes X_1 and X_2:

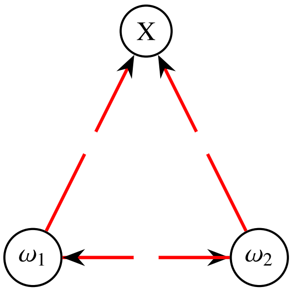


$$K_{Y(X_1,X_2)}.K^{-1}_{(X_1X_2)^2}.K_{(X_1X_2)Y} = 1 - \frac{var(Y - \Delta E)}{var(Y)} = 1 - \frac{var(Y - (E - mean(Y)))}{var(Y)} = 0.4564833$$

We notice a medium causal link $K_{Y(X_1,X_2)}.K^{-1}_{(X_1X_2)^2}.K_{(X_1X_2)Y} = 0.4564833$ with correlation weights $K_{Y(X_1,X_2)} = (0.34, 0.62)$ weak and strong. Using examples, we will show in what follows how to differentiate the notion of causation and correlation.

6 Problem:Multiple causation of two causes acting on a single variable computed from correlations

In what follows, we will consider a set of two causes $\Omega = \{\omega_1, \omega_2\}$ acting on a variable X as follows:



To this graph we attribute a matrix of correlations of the causes K_{Ω^2} and a weight vector of correlations $K_{X,\Omega}$ between the causes Ω and the variable X:

$K_{\Omega^2} = \begin{pmatrix} 1 & \rho_{\omega_1\omega_2} \\ \rho_{\omega_1\omega_2} & 1 \end{pmatrix}$ and $K_{X,\Omega} = (\rho_{\omega_1 X}, \rho_{\omega_2 X})$

Then we will present a field of correlations $K_{X,\Omega} = (\rho_{\omega_1 X}, \rho_{\omega_2 X})$ for which there is a strong causation:

$$0.6 \leq K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} < 0.799$$

We will also show the representation for a weak causation:

$$0.2 \leq K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega,X} < 0.399$$

For correlation's field $K_{X,\Omega} = (\rho_{\omega_1 X}, \rho_{\omega_2 X})$, we select correlation pairs to expose the following situations:

1. A pair of **strong correlations** between the causes Ω and the variable X that implies a **strong causation** between the causes and the variable.
2. A pair of **weak correlations** between the causes Ω and the variable X that implies a **strong causation** between the causes and the variable.
3. A pair of **strong correlations** between the causes Ω and the variable X that implies an **weak causation** between the causes and the variable.
4. A pair of **weak correlations** between the causes Ω and the variable X that implies an **weak causation** between the causes and variable.

7 Strong correlation, weak correlation and strong causation between two causes and a single variable

In what follows we will consider the matrix of causes K_{Ω^2}:

$$K_{\Omega^2} = \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}$$

From the previous matrix, we will now represent the pairs of correlations $K_{X,\Omega}$ having a strong causation $0.6 \leq K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} \leq 0.799$:

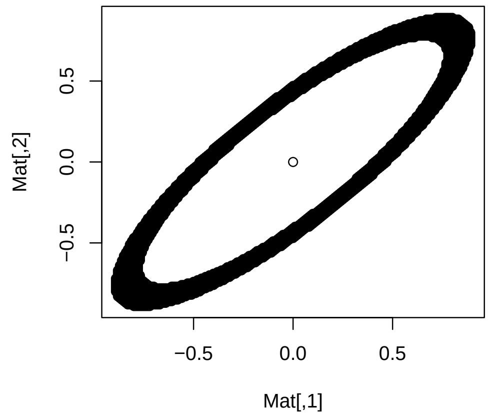


Figure 1: Pairs of correlations $K_{X,\Omega}$ having a strong causation $0.6 \leq K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} \leq 0.799$

From this graph we will select two points:$K_{X,\Omega} = (0.73, 0.77)$ and $K_{X,\Omega} = (0.18, -0.25)$. We will compute the square of multiple correlation $K_{X\Omega}.K_{\Omega^2}^{-1}.K_{X\Omega}$ for the two points:

$$K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} = (0.73, 0.77) \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}^{-1} . \begin{pmatrix} 0.73 \\ 0.77 \end{pmatrix} = 0.6134414$$

$$K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} = (0.18, -0.25) \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}^{-1} . \begin{pmatrix} 0.18 \\ -0.25 \end{pmatrix} = 0.6176577$$

We can therefore describe two situations:

1. A pair of **strong correlations** between the causes and the variable that implies a **strong causation** between the causes and the variable.
2. A pair of **weak correlations** between the causes and the variable that implies a **strong causation** between the causes and the variable.

8 Strong correlation, weak correlation and weak causation between two causes and a single variable

In what follows we will consider the same matrix of causes K_{Ω^2}:

$$K_{\Omega^2} = \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}$$

From the previous matrix, we will now represent the pairs of correlations $K_{X,\Omega}$ having a weak causation $0.2 \le K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} \le 0.399$:

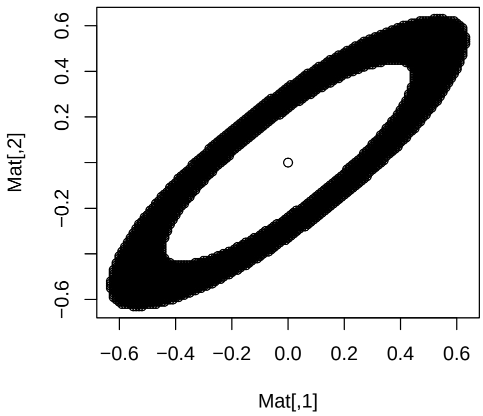


Figure 2: Pairs of correlations $K_{X,\Omega}$ having a weak causation $0.2 \le K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} < 0.399$

From this graph we will select two points:$K_{X,\Omega} = (0.61, 0.6)$ and $K_{X,\Omega} = (0.22, -0.12)$. We will compute the square of multiple correlation $K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X}$ for the two points:

$$K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} = (0.61, 0.6) \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}^{-1} . \begin{pmatrix} 0.61 \\ 0.6 \end{pmatrix} = 0.396036$$

$$K_{X\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} = (0.22, -0.12) \begin{pmatrix} 1 & 0.85 \\ 0.85 & 1 \end{pmatrix}^{-1} . \begin{pmatrix} 0.22 \\ -0.12 \end{pmatrix} = 0.388036$$

We can therefore describe two situations:

1. A pair of **strong correlations** between the causes and the variable that implies an **weak causation** between the causes and the variable.
2. A pair of **weak correlations** between the causes and the variable that implies an **weak causation** between the causes and the variable.

9 Conclusion

In this paper, we have shown mathematically the steps to follow to obtain a relationship relating the notion of causality and correlation. We have exposed the impacted and unimpacted signals from the Seatbelt matrix. From this matrix, we computed the magnitude of Gaussian causal link existing between the causes and the response variable.

Using the example of two causes acting on a variable, we have illustrated the various scenarios that may arise:

1. A pair of **strong correlations** between the causes Ω and the variable X that implies a **strong causation** between the causes and the variable.
2. A pair of **weak correlations** between the causes Ω and the variable X that implies a **strong causation** between the causes and the variable.
3. A pair of **strong correlations** between the causes Ω and the variable X that implies an **weak causation** between the causes and the variable.
4. A pair of **weak correlations** between the causes Ω and the variable X that implies an **weak causation** between the causes and variable.

[1]Optimal stastical decisions. Author: Morris H.DeGroot. Copyright 1970-2004 John Wiley and sons.

[2]Matrix Analysis. Author: Roger A.Horn and Charles R.Johnson. Copyright 2012, Cambridge university press.