Causation without correlations for the Gaussian signals

Ait-taleb Nabil *

Abstract

In this paper, we will show in a Gaussian context what to do to obtain a causal relationship between an output variable and three input variables without obtaining any correlation between the output variable and the input variables. In a context of Gaussian signals, this paper will show the following situation: Causation without correlations for Gaussian signals.

*Corresponding author: nabiltravail1982@gmail.com

1 Introduction

In this paper, we will show for Gaussian signals what to do to obtain the causality without correlations. We will place ourselves in the situation where we have three input variables $X_1X_2X_3$ and one output variable X_4. This method contains two parts:

1. Projection of a positive non-semi-definite matrix onto the subspace of positive semi-definite matrices (surface of the cone of positive semi-definite matrices SDP). This surface corresponds to the geometric domain containing the deterministic causation relationships. Concerning this projection, I advise you to read the paper "Computing Nearest correlation matrix: A problem from finance" by Nicholas Higham [4] page 9. From the paper [5] page 3, we can deduce that when we are onto the surface of the cone of the SDP matrices, we have a quadratic form $K_{X,\Omega}.K_{\Omega^2}^{-1}.K_{\Omega X} = 1$ which implies $Var(X - E[X|\Omega]) = 0$ and which means that we have the causal relationship $X = E[X|\Omega]$.

2. The second part concerns the inference of data from the library (mvtnorm) to be downloaded to the R software. I then expose, from a matrix projected on the SDP cone, a situation where we have a deterministic causation between the input variables $X_1X_2X_3$ and the output variable X_4 without having any correlations between the input variables $X_1X_2X_3$ and the output variable X_4.

In this paper, we will explain the method which makes it possible to obtain a causal relationship without correlations. The paper ends with a simulation with the R software in order to expose the situation: causation without correlation for Gaussian signals.

2 Causal relationship obtained from a correlation matrix

We will consider three input signals (variables) X_1, X_2, X_3 and one output signal X_4 (the response) below.

We will show below the steps to follow to obtain a causal relationship between the output X_4 and the inputs X_1, X_2, X_3 without having any correlations between the inputs and the output.

Method

1. Choose a symmetric matrix M located outside the positive semi-definite matrix cone with $m_{ii} = 1$ and $-1 < m_{ij} < 1$.

   The rows and columns correspond to the variables in order at $X1, X2, X3, X4$.

   The last row and column X4 of the matrix M have a small value. In our example we chose the following matrix:

$$M = \begin{pmatrix} 1.00 & -0.61 & 0.60 & 0.01 \\ -0.61 & 1.00 & 0.65 & 0.02 \\ 0.60 & 0.65 & 1.00 & 0.03 \\ 0.01 & 0.02 & 0.03 & 1.00 \end{pmatrix}$$

2. We will now choose a weight vector of diagonal elements: $(10, 20, 30, 40)$ to build the matrix A.

$$A = \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 40 \end{pmatrix} \begin{pmatrix} 1.00 & -0.61 & 0.60 & 0.01 \\ -0.61 & 1.00 & 0.65 & 0.02 \\ 0.60 & 0.65 & 1.00 & 0.03 \\ 0.01 & 0.02 & 0.03 & 1.00 \end{pmatrix} \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 40 \end{pmatrix}$$

$$A = \begin{pmatrix} 100 & -122 & 180 & 4 \\ -122 & 400 & 390 & 16 \\ 180 & 390 & 900 & 36 \\ 4 & 16 & 36 & 1600 \end{pmatrix}$$

   Note that the eigenvalues are as follows: $\vec{\lambda} = (1603.25899, 1120.04254, 336.70282, -60.00435)$. So this symmetric matrix is not positive semi-definite. We now arrive at the stage where we must obtain a variance-covariance matrix which means a positive semi-definite symmetric matrix A_+.

3. We use now the projection $A_+ = P_S(A)$ onto the cone of semi definite positive matrices of Nicholas Higham paper [4] page 9 to obtain a singular semi-definite positive matrix: $A_+ = Q.diag(max(\lambda_i, 0)).Q^T$ and $P_S(A) = W^{-1/2}((W^{1/2}.A.W^{1/2})_+).W^{-1/2}$.

   We will use as weight matrix W like the identity matrix I.

4. Now we will choose a random mean vector $\vec{\mu} = (10, 20, 30, 40)$, and use the $rmvnorm()$ function to infer 1000 data with a fixed mean vector $\vec{\mu}$ and a fixed variance covariance matrix A_+: $Data = rmvnorm(1000, \vec{\mu}, A_+)$. On the boundary of the cone of positive semi-definite matrices, the matrix $A+$ is singular and therefore has the quadratic form:

$$(A_+)_{X_4,(X_1X_2X_3)} . (A_+)^{-1}_{(X_1X_2X_3)^2} . (A_+)_{(X_1X_2X_3),X_4} = 1$$

, using paper [5]:

$$X_4 = E[X_4|X_1X_2X_3] = \beta_{14}.X_1 + \beta_{24}X_2 + \beta_{34}.X_3 + \beta_4$$

In the R code in the appendix, we will test E[.] and X4[.] to see if the values are the same. As this will be the case, we will be able to say that the causal relationship will be perfect and this without the presence of correlations between the inputs X_1 X_2 X_3 and the output X_4.

In what follows, we will represent, from R software (see appendix), the input signals $X_1X_2X_3$ and the output signal X_4.

3 Input signals $X_1X_2X_3$ and output signal X_4

Below we will represent the signals at the input $X_1X_2X_3$ and the output signal X_4:

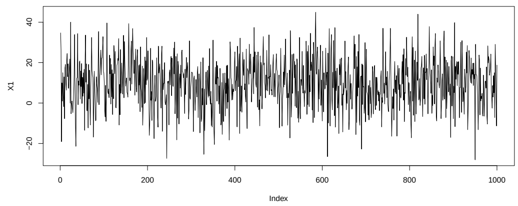

Figure 1: input signal X_1

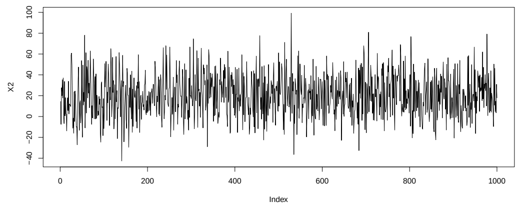

Figure 2: input signal X_2

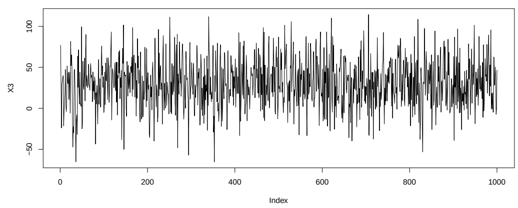

Figure 3: Input signal X_3

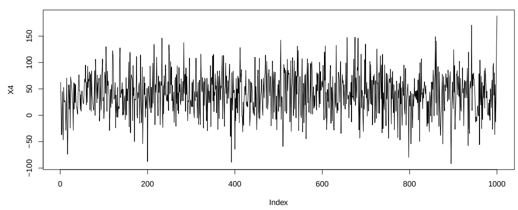

Figure 4: Output signal X_4

A Simulation from R software

```
> M=diag(4)
> M[1,2]=-0.61
> M[2,1]=-0.61
> M[1,3]=0.6
> M[3,1]=0.6
> M[1,4]=0.01
> M[4,1]=0.01
> M[2,3]=0.65
> M[3,2]=0.65
> M[2,4]=0.02
> M[4,2]=0.02
> M[3,4]=0.03
> M[4,3]=0.03
> library(mvtnorm)
> P=diag(4)
> diag(P)=c(10,20,30,40)
> M=P
> Q=eigen(M)[[2]]
> M1=diag(4)
> diag(M1)=eigen(M)[[1]]
> M1[4,4]=0
> Aplus=Q
> D=rmvnorm(1000,c(10,20,30,40),Aplus)
> X1=D[,1]
> X2=D[,2]
> X3=D[,3]
> X4=D[,4]
> MV=var(D)
> tV=MV[1:3,4]
> B=tV
```

```
> C=mean(X4)-B[1]*mean(X1)-B[2]*mean(X2)-B[3]*mean(X3)

> E=B[1]*X1+B[2]*X2+B[3]*X3+C

> E[1]

[1] 56.73746

> X4[1]

[1] 56.73746

> E[14]

[1] 7.352529

> X4[14]

[1] 7.352529

> E[105]

[1] -3.265968

> X4[105]

[1] -3.265968

etc

> cor(D)

X1 1.000000000 -0.44096948 0.4579423 -0.001413494

X2 -0.440969482 1.00000000 0.5959369 0.009629230

X3 0.457942308 0.59593688 1.0000000 0.011557804

X4 -0.001413494 0.00962923 0.0115578 1.000000000
```

Note that there are no correlations between the output X_4 and the inputs X_1, X_2 and X_3: $K_{X4,(X_1,X_2,X_3)} = (-0.001413494, 0.00962923, 0.0115578$, however the signals X_1, X_2, X_3 predicted all the values of signal X_4. In this situation, there are no correlations but there is definitely a causal relationship:

$$X_4 = E[X_4|X_1X_2X_3] = \beta_{14}.X_1 + \beta_{24}X_2 + \beta_{34}.X_3 + \beta_4$$

$$X_4 = -891,6246.X_1 - 576,9075.X_2 + 401,5839.X_3 + 8446.877$$

[1]Elements of information theory. Author: Thomas M.Cover and Joy A.Thomas. Copyright 1991 John Wiley and sons.

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

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

[4]Computing the nearest correlation Matrix-A problem from finance. Author: Nicholas Higham.copyright 2002, The university of Manchester

[5]Causal effect vector and multiple correlation. Author Nabil Ait-Taleb. copyright 2024