A Note on Causation Versus Correlation in an Extreme Situation

2 authors, including:

X. San Liang
Nanjing Institute of Meteorology
114 PUBLICATIONS **1,548** CITATIONS

Some of the authors of this publication are also working on these related projects:

Atmospheric dynamics View project

Canonical Transfers in the Ocean View project

A Note on Causation Versus Correlation in an Extreme Situation

X. San Liang [1],* and Xiu-Qun Yang [2]

1   Nanjing Institute of Meteorology, 219 Ningliu Blvd, Nanjing 210044, China
2   School of Atmospheric Sciences, Nanjing University, 163 Xianlin Avenue, Nanjing 210023, China; xqyang@nju.edu.cn
*   Correspondence: x.san.liang@gmail.com

Abstract: Recently, it has been shown that the information flow and causality between two time series can be inferred in a rigorous and quantitative sense, and, besides, the resulting causality can be normalized. A corollary that follows is, in the linear limit, causation implies correlation, while correlation does not imply causation. Now suppose there is an event A taking a harmonic form (sine/cosine), and it generates through some process another event B so that B always lags A by a phase of $\pi/2$. Here the causality is obviously seen, while by computation the correlation is, however, zero. This apparent contradiction is rooted in the fact that a harmonic system always leaves a single point on the Poincaré section; it does not add information. That is to say, though the absolute information flow from A to B is zero, i.e., $T_{A \to B} = 0$, the total information increase of B is also zero, so the normalized $T_{A \to B}$, denoted as $\tau_{A \to B}$, takes the form of $\frac{0}{0}$. By slightly perturbing the system with some noise, solving a stochastic differential equation, and letting the perturbation go to zero, it can be shown that $\tau_{A \to B}$ approaches 100%, just as one would have expected.

Keywords: causality; time series; information flow; correlation

1. A Review of the Rigorous Information Flow-Based Causality Analysis

Causal inference is a fundamental problem in scientific research. Recently it has been shown that the problem can be recast into the framework of information flow, another fundamental notion in general physics which has wide applications in different disciplines (see [1]), and hence can be put on a rigorous footing. The causality between two time series can then be analyzed in a quantitative sense, and, besides, the resulting formula is very concise in form. In the linear limit, it involves only the usual statistics namely sample covariances [2], making the important and otherwise difficult problem an easy task.

To briefly review the theory, consider a two-dimensional continuous-time stochastic system for state variables $\mathbf{x} = (x_1, x_2)$

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) + \mathbf{B}(\mathbf{x}, t)\dot{\mathbf{w}}, \tag{1}$$

where $\mathbf{F} = (F_1, F_2)$ may be arbitrary nonlinear functions of $\mathbf{x}$ and t, $\dot{\mathbf{w}}$ is a vector of white noise, and $\mathbf{B} = (b_{ij})$ is the matrix of perturbation amplitudes which may also be any functions of $\mathbf{x}$ and t. Here we adopt the convention in physics and do not distinguish deterministic and random variables; in probability theory, they are ususally distinguished with capital and lower-case symbols. Assume that $\mathbf{F}$ and $\mathbf{B}$ are both differentiable with respect to $\mathbf{x}$ and t. Then the information flow from x_2 to x_1 (in nats per unit time) can be explicitly found in a closed form [3] (the multiple-dimensional case is referred to [1]):

$$T_{2 \to 1} = -E\left[\frac{1}{\rho_1} \frac{\partial(F_1 \rho_1)}{\partial x_1}\right] + \frac{1}{2}E\left[\frac{1}{\rho_1} \frac{\partial^2 g_{11} \rho_1}{\partial x_1^2}\right], \tag{2}$$

where E stands for mathematical expectation, and $g_{ii} = \sum_{k=1}^{n} b_{ik} b_{ik}$, $\rho_i = \rho_i(x_i)$ is the marginal probability density function (pdf) of x_i. The rate of information flowing from x_1 to x_2 can be obtained by switching the indices. If $T_{j \to i} = 0$, then x_j is not causal to x_i; otherwise it is causal, and the absolute value measures the magnitude of the causality from x_j to x_i. For discrete-time mappings, the information flow is in much more complicated a form; see [1].

In the case with only two time series (no dynamical system is given) X_1 and X_2, under the assumption of a linear model with additive noise, the maximum likelihood estimator (MLE) of the rate of information flowing from X_2 to X_1 is [2]

$$\hat{T}_{2 \to 1} = \frac{C_{11} C_{12} C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11} C_{12}^2},$$
(3)

where C_{ij} is the sample covariance between X_i and X_j, and $C_{i,dj}$ the sample covariance between X_i and a series derived from X_j using the Euler forward differencing scheme (also see the Euler–Maruyama scheme in [4]): $\dot{X}_{j,n} = (X_{j,n+k} - X_{j,n})/(k\Delta t)$, with $k \geq 1$ some integer. Note that Equation (3) is rather concise in form; it only involves the common statistics, i.e., sample covariances. In other words, a combination of some sample convariances will give a quantiative measure of the causality between the time series. This makes causality analysis, which otherwise would be complicated with the classical empirical/half-empirical methods, very easy. Nonetheless, note that Equation (3) cannot replace (1); it is just the mle of the latter. Statistical significance test must be performed before a causal inference is made based on the computed $T_{2 \to 1}$. For details, refer to [2].

Considering the long-standing debate ever since Berkeley (1710) [5] over correlation versus causation, we may rewrite (3) in terms of linear correlation coefficients, which immediately implies [2]:

Causation implies correlation, but correlation does not imply causation.

The above formalism has been validated with many benchmark systems (e.g., [1]) such as baker transformation, Hénon map, Kaplan-Yorke map, Rössler system, etc. It also has been successfully applied to the studies of many real world problems such those in financial economics (e.g., the "Seven Dwarfs vs. a Giant" problem [6]), earth system science (e.g., the Antarctica mass balance problem [7] and the global warming problem [8]), neuroscience (e.g., the concussion problem [9]), to name but a few.

2. The Question

Now suppose we have a dynamic event A which drives another event B. The former has a harmonic form, leading the latter by a phase of $\pi/2$. That is to say, the time series, say, $\{x_A(t)\}$ and $\{x_B(t)\}$ resulting from the two, are in quadrature. Then the correlation between the two is zero. Here by zero-correlation we mean a zero integral

$$\int_{\Omega} (x_A(t) - \bar{x}_A)(x_B(t) - \bar{x}_B) dt,$$

with the integration domain Ω being one period or more periods, and the overbar being the mean over the domain. However, since A causes B, the result is apparently in contradiction to the above corollary that "causation implies correlation".

3. The Solution

The problem can be more formally stated with the harmonic system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) = \mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$
(4)

If the system is initialized with $x_1(0) = 1$, $\dot{x}_1(0) = 1$), the solution is, $x_1 = \cos t$, $x_2 = \sin t$. Thus, the population covariance $\sigma_{12} = \int_\Omega \cos t \sin t dt = 0$ (Ω is one period or many periods). This yields an information flow from x_2 to x_1:

$$T_{2\to1} = a_{12}\frac{\sigma_{12}}{\sigma_{11}} = 0. \tag{5}$$

Fundamentally the above problem arises from the fact that it is a deterministic system. In Granger causality test [10] (also see a recent reference [11]), this case has been explicitly excluded, as in such case the trajectories do not form appropriate ensembles in the sample space. For a harmonic series, it shows on a Poincaré section only one single point; so the total information does not accrue. If the total information does not change, the information flow to x_1 must also vanish. However, the vanishing information flow does not mean that there is no influence of x_2 on x_1. As we argued in Liang (2015), the so-obtained information must be normalized, just as covariance needs to be normalized into correlation, for one to assess the causal influence. Here if the normalizer is zero, $T_{2\to1}$ involves the indeterminate form $\frac{0}{0}$. We may then approach it by taking the limit. Specifically, we may approach it by enlarging the sample space slightly, i.e., by adding some stochasticity to the system, then take the limit by letting the stochastic perturbation amplitude go zero.

By Liang (2015), the normalizer for $T_{2\to1}$ is

$$Z_{2\to1} = |T_{2\to1}| + \left|\frac{dH_1^*}{dt}\right| + \left|\frac{dH_1^{\text{noise}}}{dt}\right|, \tag{6}$$

where on the right hand side, the second term is the contribution from x_1 itself, and the third term the contribution from noise. In Liang (2015), it has shown that $\frac{dH_1^*}{dt}$ is a Lyapunov exponent-like, phase-space stretching rate, and $\frac{dH_1^{\text{noise}}}{dt}$ a noise-to-signal ratio. In this problem, we do not have noise taken into account. However, in reality, noise is ubiquitous. We may hence view a deterministic system as a limit or extreme case as the amplitude of stochastic perturbation goes to zero. For this case, we add to (4) a stochastic term:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{B}\dot{\mathbf{w}}, \tag{7}$$

where $\mathbf{w}$ is a vector of standard Wiener processes. For simplicity, let the perturbation amplitude $\mathbf{B}$ a constant matrix. Further let $\mathbf{G} = \mathbf{B}\mathbf{B}^T$, with elements

$$(g_{ij}) = \sum_{k=1}^{2} b_{ik}b_{jk}.$$

Liang (2008) established that

$$\frac{dH_1^*}{dt} = a_{11} = 0, \tag{8}$$

$$\frac{dH_1^{\text{noise}}}{dt} = \frac{1}{2}\frac{g_{11}}{\sigma_{11}}. \tag{9}$$

So in this case, the normalized flow from x_2 to x_1 is

$$\tau_{2\to1} = \frac{a_{12}\frac{\sigma_{12}}{\sigma_{11}}}{\left|a_{12}\frac{\sigma_{12}}{\sigma_{11}}\right| + 0 + \left|\frac{1}{2}\frac{g_{11}}{\sigma_{11}}\right|} = \frac{-\sigma_{12}}{|\sigma_{12}| + \frac{g_{11}}{2}}. \tag{10}$$

Likewise,

$$\tau_{1\to2} = \frac{a_{21}\frac{\sigma_{12}}{\sigma_{22}}}{\left|a_{21}\frac{\sigma_{12}}{\sigma_{22}}\right| + |a_{22}| + \left|\frac{1}{2}\frac{g_{22}}{\sigma_{22}}\right|} = \frac{+\sigma_{12}}{|\sigma_{12}| + \frac{g_{22}}{2}}. \tag{11}$$

Note that $\tau_{2\to1}$ (or $\tau_{1\to2}$) may be positive or negative. In causal inference, this does not matter; we need only consider the absolute value, although the sign does carry a meaning according to the original formulation. (A positive $\tau_{2\to1}$ means x_2 causes the marginal entropy of x_1 to grow and vice versa; see [1,2] .)

Now for the stochastic equation, the covariance matrix $\boldsymbol{\Sigma}$ evolves as

$$\frac{d\boldsymbol{\Sigma}}{dt} = \mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T + \mathbf{G} \tag{12}$$

Expanding, this is

$$\frac{d}{dt}\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} -\sigma_{12} & -\sigma_{22} \\ \sigma_{11} & \sigma_{12} \end{bmatrix} + \begin{bmatrix} -\sigma_{12} & \sigma_{11} \\ -\sigma_{22} & \sigma_{12} \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix}.$$

We hence obtain the following equation set:

$$\frac{d\sigma_{11}}{dt} = -2\sigma_{12} + g_{11},$$
$$\frac{d\sigma_{12}}{dt} = -\sigma_{22} + \sigma_{11} + g_{12},$$
$$\frac{d\sigma_{22}}{dt} = 2\sigma_{12} + g_{22}.$$

Solving, we get

$$\frac{d^2\sigma_{12}}{dt^2} = -4\sigma_{12} + (g_{11} - g_{22} + g_{12}).$$

So the solution is

$$\sigma_{12} = C_1 \cos 2t + C_2 \sin 2t + \frac{1}{2}(g_{11} - g_{22} + g_{12})t^2.$$

If $\sigma_{12}(0) = 0$, $\dot\sigma_{12}(0) = 0$, then the integration constants $C_1 = C_2 = 0$. So

$$\tau_{2\to1} = \frac{-\sigma_{12}}{|\sigma_{12}| + \frac{1}{2}g_{11}} = \frac{-1}{1 + \frac{g_{11}}{(g_{11}-g_{22}+g_{12})t^2}}$$

Two cases are distinguished:

Case I $g_{12} - g_{22} = \text{const} \neq 0$.

$$\lim_{g_{11}\to0} \tau_{2\to1} = -1.$$

Case II $g_{12} - g_{22} = 0$.

$$\lim_{g_{11}\to0} \tau_{2\to1} = \frac{-1}{1 + 1/t^2}.$$

As t goes to infinity, $\tau_{2\to1}$ also approaches -1.

If initially there exists some covariance, say, $\sigma_{12}(0) = c$, then $C_1 = c$, and hence

$$\tau_{2\to1} = \frac{-1}{1 + \frac{g_{11}}{2c\cos 2t + (g_{11}-g_{22}+g_{12})t^2}}.$$

In this case, as $g_{11} \to 0$, we always have $\tau_{2\to 1} \to -1$. Either way, the relative information flow $\tau_{2\to 1}$ approaches -1 in the limit of deterministic system.

In the other direction, we now need to consider the uncertainty growth of x_2 and hence perturb g_{22}. Repeating the above procedure, when $\sigma_{12}(0) = \dot{\sigma}_{12}(0) = 0$, the normalized information flow is

$$\tau_{1\to 2} = \frac{+\sigma_{12}}{|\sigma_{12}| + \frac{1}{2}g_{22}} = \frac{+1}{1 + \frac{g_{22}}{(g_{11}-g_{22}+g_{12})t^2}}.$$

If $g_{11} + g_{12} = \text{const} \neq 0$, then

$$\lim_{g_{22}\to 0} \tau_{1\to 2} = +1;$$

else $(g_{11} + g_{12} = 0)$

$$\lim_{g_{22}\to 0} \tau_{1\to 2} = \frac{1}{1 - 1/t^2}$$

which approaches to 1 for enough long time $(t \to \infty)$. On the other hand, if initially there exists some covariance such that $\sigma_{12}(0) = c$ then

$$\tau_{12} = \frac{1}{1 + \frac{g_{22}}{2c\cos 2t + (g_{11}-g_{22}+g_{12})t^2}}$$

which implies

$$\lim_{g_{22}\to 0} \tau_{1\to 2} = 1.$$

This is indeed what we expect. So even for this extreme case, there is no contradiction at all for causal inference using information flow.

4. Discussion

To summarize, a recent rigorously formulated causality analysis asserts that, in the linear limit, causation implies correlation, while correlation does not necessarily mean causation. In this short note, an extreme case which apparently contradicts to the assertion is examined. In this case an event x_1 takes a harmonic form (sine/cosine), and generates through some process another event x_2 so that x_2 is always out of phase with x_1, i.e., lag x_1 by $\pi/2$. Obviously x_1 causes x_2, but by computation the correlation between x_1 and x_2 is zero. In this study we show that this is an extreme case, with only one point in the phase space and hence the problem becomes singular. We re-examine the problem by enlarging the ensemble space slightly through adding some noise. A stochastic differential equation is then solved for the corresponding covariances, which allows us to obtaint the information flows for the perturbed system. Then as the noisy perturbation goes to zero, the normalized information flow rate from x_1 to x_2 is established to be 100%, just as one would have expected. So actually no contradiction exists. (see [12] for how a stochastic differential equation is solved by perturbing it with noise [12].)

One thing that merits mentioning is that, here although it seems that x_1 causes x_2, actually here the normalized information flow rate from x_2 to x_1 is also 100%. That is to say, for such a harmonic system with circular cause-effect relation, it is actually impossible to differentiate causality by simply assessing which takes place first; anyhow, taking lead by $\pi/2$ is equivalent to lagging by $3\pi/2$. The moral is, for a process that is nonsequential (e.g., that in the nonsequential stochastic control systems), circular cause and consequence coexist, it is essentially impossible to distinguish a delay from an advance.

References

1. Liang, X.S. Information flow and causality as rigorous notions *ab initio*. *Phys. Rev. E* **2016**, *94*, 052201.
2. Liang, X.S. Unraveling the cause-effect relation between time series. *Phy. Rev. E* **2014**, *90*, 052150.
3. Liang, X.S. Information flow within stochastic dynamical systems. *Phys. Rev. E* **2008**, *78*, 031113.
4. Iacus, S.M. *Simulation and Inference for Stochastic Differential Equations: With R Examples*; Springer: New York, NY, USA, 2008.
5. Berkeley, G. *A Treatise Concerning the Principles of Human Knowledge*; originally published in 1710; Hackett Publishing Company, Inc.: Indianapolis, IN, USA, 1982.
6. Liang, X.S. Normalizing the causality between time series. *Phys. Rev. E* **2015**, *92*, 022126.
7. Vannitsem, S.; Dalaiden, Q.; Goosse, H. Testing for dynamical dependence—Application to the surface mass balance over Antarctica. *Geophys. Res. Lett.* **2019**, doi:10.1029/2019GL084329.
8. Stips, A.; Macias, D.; Coughlan, C.; Garcia-Gorriz, E.; Liang, X.S. On the causal structure between CO_2 and global temperature. *Sci. Rep.* **2016**, *6*, 21691, doi:10.1038/srep21691.
9. Hristopulos, D.T.; Babul, A.; Babul, S.; Brucar, L.R.; Virji-Babul, N. Disrupted information flow in resting-state in adolescents with sports related concussion. *Front. Hum. Neurosci.* **2019**, *13*, 419, doi:10.3389/fnhum2019.00419.
10. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.
11. Contreras-Reyes, J.E., Hernández-Santoro, C. Assessing Granger-causality in the southern Humboldt current ecosystem using cross-spectral methods. *Entropy* **2020**, *22*, 1071.
12. Argyris, J.; Andreadis, I.; Pavlos, G.; Athanasiou, M. The influence of noise on the correlation dimension of chaotic attractors. *Chaos Solit. Fract.* **1998**, *9*, 343–361.