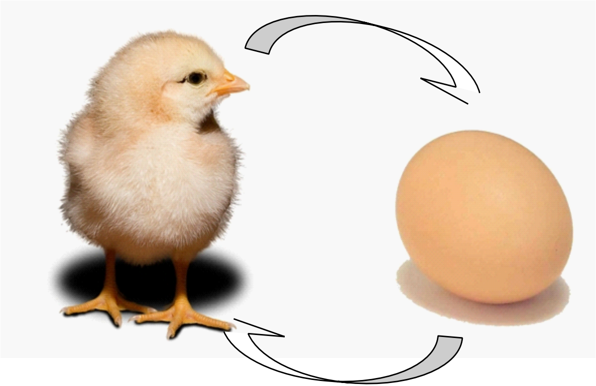
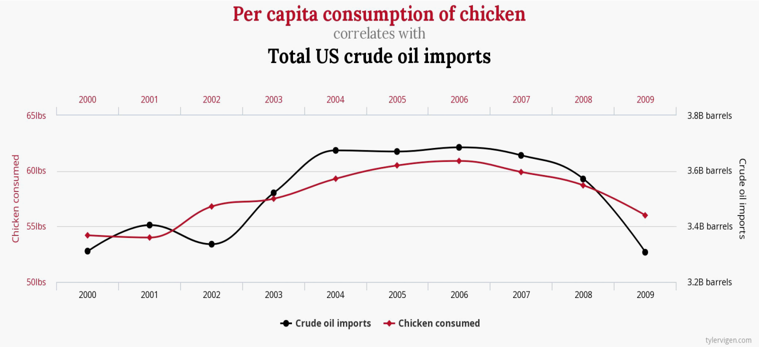
A unified theory and method for detection of causal relationships

Zs. Benkő, Á, Zlatniczki, D. Fabó, D. Erőss, A. Telcs, Z. Somogyvári
Wigner RCP,OKITI, BME
telcs.andras@wigner.hu

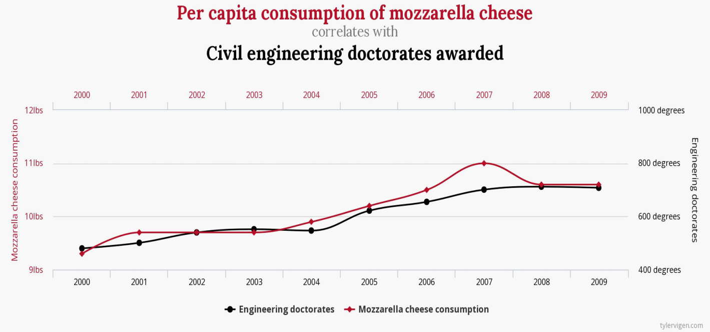
September 22, 2020

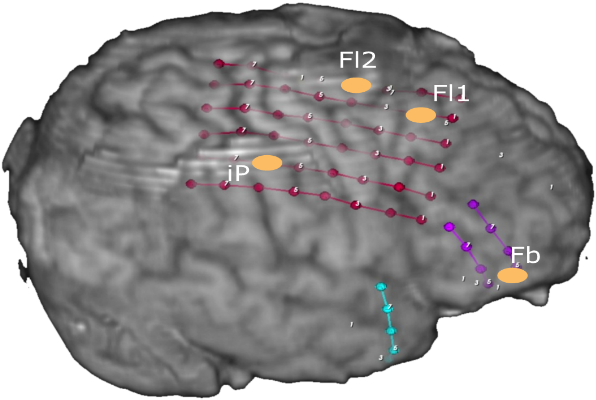
A puzzle of 2500 years

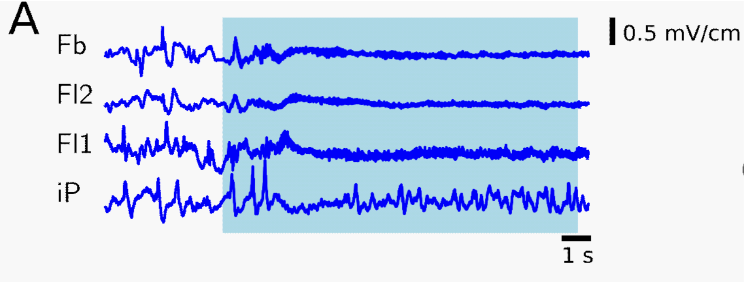




Per capita consumption of chicken
correlates with
Total US crude oil imports
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
65lbs
60lbs
55lbs
50lbs
Chicken consumed
3.8B barrels
3.6B barrels
3.4B barrels
3.2B barrels
Crude oil imports
Crude oil imports
Chicken consumed
tylervigen.com


Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
12lbs
11lbs
10lbs
9lbs
Mozzarella cheese consumption
1000 degrees
800 degrees
600 degrees
400 degrees
Engineering doctorates
Engineering doctorates
Mozzarella cheese consumption
tylervigen.com


FI2
FI1
iP
Fb


A
Fb
FI2
FI1
iP
0.5 mV/cm
1 s

1. basic questions
2. result on deterministic dynamic systems
3. causality between stochastic variables
4. causality between stochastic dynamic systems

Causation between events

$$A \rightarrow B$$

common cause C

$$\begin{aligned} C &\rightarrow A \\ C &\rightarrow B \end{aligned}$$

Causation between events

Example

A =sun burn
B =had ice cream
$A \rightarrow ?B$ or $B \rightarrow ?A$ or $A \leftrightarrow ?B$

No, there is a third hidden common cause $C =$ suny, hot day

$$\begin{aligned} C &\rightarrow A \\ C &\rightarrow B \end{aligned}$$

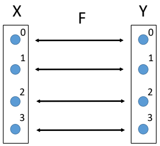
$$A \perp B | C$$

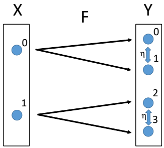
inference...

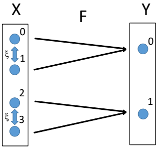
Causation between random variables

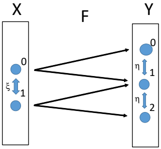
When can we say that X causes Y? If $X = I_A, Y = I_B$ we know. But in general?
Alternatives:
X determines Y ,for X values there is only one Y value,, there is an f: $Y = f(X)$.
If X "changes" then Y "changes". For all Y values there is only one X value.there is a g: $X = g(Y)$.
X and Y mutually determines each other. $Y = f(X)$ and $X = f^{-1}(Y)$
X influences Y for some changes of X, Y changes and sometimes Y changes without change of X.Are we satisfied?

Causation between random variables









Causation between random variables

Very tempting to say X causes Y if any change in X causes change in Y. (first two cases) But excludes the simple possible case when X is observed with noise or little change in X has no effect on Y. (discussed latare again)
In general

Definition

X cause Y, if there is a function f and r.v. η s.t. $X \perp \eta$ and

$$Y = f(X, \eta).$$

Denoted by $X \rightarrow Y$.

Reichenbach's common cause principle 1956

(informal recall)
If A and B correlated then there is a third C causing both and thay are conditionally independent given C

Common cause

Definition

Z is common cause of X and Y if

$$X \perp Y | Z.$$

Lemma

If X and Y has joint distributon they have a common cause.

Common cause

Definition

Z is common cause of X and Y if

$$X \perp Y | Z.$$

Lemma

If X, Y and Z satisfies

$$X \perp Y | Z.$$

then Z causes X and Y.

Common cause trivial cases

Remark

Assume Z is common cause of X and Y.
If $X \perp Y$ then Z is trivial, $Z \equiv c$.
Z bijective image of X iff $X \rightarrow Y$
Z bijective image of Y iff $Y \rightarrow X$
Z bijective image of X and Y iff X bijective image of Y that is $X \leftrightarrow Y$.

Test of local independence

Remark

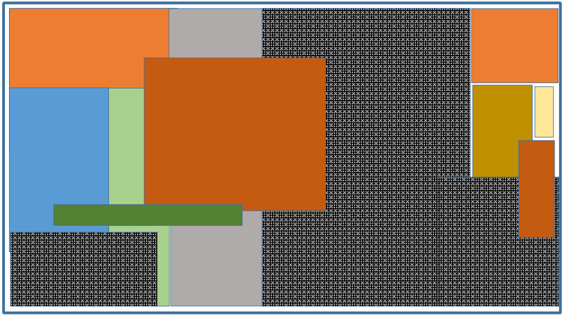
From now on we assume that X and Y are discrete RW.

If we have a sample $\{(x_i, y_i)\}_{i=1}^n$ from the joint (X, Y) let us find the common cause. That is a $Z:\ X \perp Y|Z$. We have the contingency table of the sample.$\{n_{i,j}\}$ with marginals r_i, c_j for Y and X.

29	5	7	17
54	10	19	5
37	15	12	10
Y/X	30	38	32

Common information

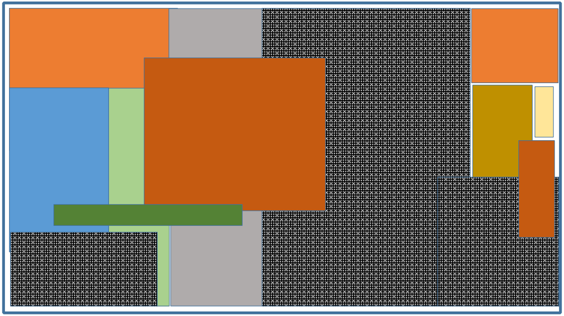
Let $\mathcal{Q}_{X,Y}$ a submatrix cover of the $Dom(X) \times Dom(Y)$:



A cover is a common cause if for all submatrix $z \in \mathcal{Q}_{X,Y}$

$$X \perp Y \mid_z .$$

Common information



The $\mathcal{Q}_{X,Y}$ cover corresponds to the screening/ common cause variable Z.The $Z = z$ atoms to the submatrices.

Common information

Wyner 1975 common information

$$W = \arg\min \left\{ I\left(X, Y : W\right) : W : X \perp Y | W \right\}$$

$$C\left(X, Y\right) = I\left(X, Y : W\right)$$

Wyner's' common information

$$W = \arg\min \{ I(X, Y : W) : W : X \perp Y | W \}$$

by definition $H(W, X, Y) = H(X, Y)$ consequently

$$H(W|X, Y) = 0.$$

For any (x, y) there is a unique w, that is there is an f s.t.

$$W = f(X, Y),$$

can be "reconstructed" from X and Y.
No overlaps.

Common information

Let $\mathcal{P}_{X,Y}$ a submatrix partition of the $Dom(X) \times Dom(Y)$:

A partition is a common cause if for all $z \in \mathcal{P}_{X,Y}$

$$X \perp Y \,\big|_{z}.$$

Common information

For partitions $H(\mathcal{P}_{X,Y}|X,Y)=0$. The $\mathcal{P}_{X,Y}$ corresponds to the common cause/screening variable Z,$Z=z$ correspond to atoms, to the rectangles.

Common information

Common information/variable Gács-Körner

$$C_{GK}(X,Y) = \max_{\substack{V:V=f(X)\\ V=g(Y)}} I(V:X,Y)$$

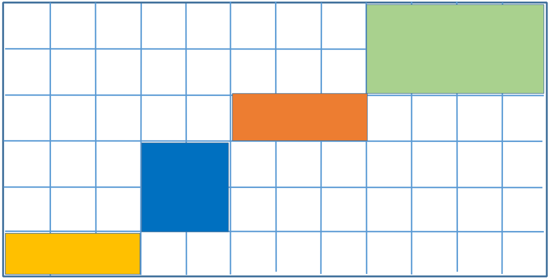
$$Z_{GK} = \arg\max \{I(V:X,Y); V:V=f(X), V=g(Y)\}$$

Theorem

$$C_{GK}(X,Y) \leq I(X,Y) \leq C_W(X,Y)$$

and $=$ *holds iff* $X=(U,Q)$, $Y=(V,Q)$ *,where U,V,Q are independent.*

Common information - Gács-Körner



Search for common cause

Use a variant if Wyner's method (very tricky) or a heuristic one.
We define a goal function to be minimized.
Let $MI(\mathcal{P})$, a measure of local independence of submatrices in $\mathcal{P}$ (=zero for perfect independence, c.f. χ^2 statistics)
$C(\mathcal{P})$ model complexity: the number of blocks or Shannon entropy of thecover/ partition elements ,

$$L(\mathcal{P}) = \alpha MI(\mathcal{P}) + \beta C(\mathcal{P})$$

where α, β are regularizing meta parameters. Global minimum is not guaranteed.

Search for common cause

Basic seach strategy. Call a C submatrix block if $X \perp Y|_C$.

Remark

If C is a a block, all sub-matrices of it are also blocks

algorithmic steps:

- find all 2x2 blocks
- start from a random 2x2 block and extend it to a random direction if possible to obtain a 2x3 or 2x3 block
- continue the extension until it is possible each step in a random direction
- if still there is an unused 2x2 block go to step 2. else stop.
- select maximal blocks (not contained by any other).

Search for common cause

Generate many partition candidate with the above randomized algorithm, calculate the goal function and select the best. Meta parameters of the algorithm to be adjusted: binning parameter of $[0,1]^2$ for the contingency table (do binning by with or quantile values for X and Y) α, β the regularization parameters.

Causation between stochastic processe

Assume X_t, Y_t are stationary stochastic processes, (f.s.s. 1-order stationary Markov chains)

$$\begin{aligned} X_{n+1} &= a\left(X_n, \xi_{n+1}\right) \\ Y_{n+1} &= b(X_n, Y_n, \eta_{n+1}) \end{aligned}$$

where ξ is i.i.d uniform and independent of X_0 and similarly η is i.i.d. and independent of X_0, Y_0.

Common cause of stochastic processe

Assume X_t, Y_t are stationary stochastic processes, Z_t is common cause if

$$
\begin{aligned}
X_{n+1} &= a\left(X_n, Z_{n,}\xi_{n+1}\right) \\
Y_{n+1} &= b(Y_n, Z_n, \eta_{n+1}) \\
Z_{n+1} &= c\left(Z_n, \zeta_{n+1}\right)
\end{aligned}
$$

and

$$X_{n+1} \perp Y_{n+1} | X_n, Y_n, Z_n$$

Find the common cause

one can use the $(X_{n+1}, X_n), (Y_{n+1}, Y_n)$ variables and find block partition of the joint or ...

Find the common cause

observe that

$$Z_{n+1} = c\left(Z_n, \zeta_{n+1}\right)$$

defines a 1-order Markov chain and if it is CC then

$$\begin{aligned} X_{n+1} &= a\left(X_n, Z_{n,}\xi_{n+1}\right) \\ Y_{n+1} &= b(Y_n, Z_n, \eta_{n+1}) \end{aligned}$$

means that (X_n, Z_n) and (Y_n, Z_n) are also 1-order Markov chains.

Find the common cause

$$Z_{n+1} = c\left(Z_n, \zeta_{n+1}\right)$$

$$\begin{aligned} X_{n+1} &= a\left(X_n, Z_{n,} \xi_{n+1}\right) \\ Y_{n+1} &= b(Y_n, Z_n, \eta_{n+1}) \end{aligned}$$

observe that with

$$X_{n+1} \perp Y_{n+1} | X_n, Y_n, Z_n$$

it is the standard specification of the hidden Markov Model (HMM) with the hidden Z_n MC.