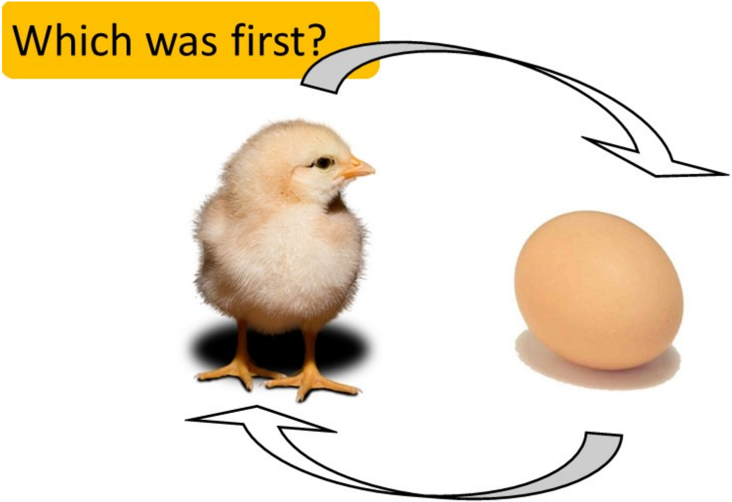
Unified theory of temporal causal discover

Antal Jakovác, András Telcs
HUN-REN Wigner RCP

November 21, 2024

The question



Philosophy

Our causality notion is based on time.

We observe, recognize time through irreversibility.

Philosophy

Our causality notion is based on time.
We observe, recognize time through irreversibility.
Irreversibility is created by randomness.

Philosophy

Our causality notion is based on time.
We observe, recognize time through irreversibility.
Irreversibility is created by randomness.

In case of perfect observation of a deterministic system, without intervention it is impossible to find the direction of driving the direction of causation.

Philosophy

Our causality notion is based on time.
We observe, recognize time through irreversibility.
Irreversibility is created by randomness.

In case of perfect observation of a deterministic system, without intervention it is impossible to find the direction of driving the direction of causation. In case of continuous space there is no perfect observation at all.

Philosophy

Our causality notion is based on time.
We observe, recognize time through irreversibility.
Irreversibility is created by randomness.

In case of perfect observation of a deterministic system, without intervention it is impossible to find the direction of driving the direction of causation. In case of continuous space there is no perfect observation at all.

Life

Survival needs some kind of inteligence,
first of all detection of causal relations.

Life

Survival needs some kind of inteligence,
first of all detection of causal relations.

The question

We observe two time series: $X = \{x_i\}, Y = \{y_i\}$ that is observation of two dynamic systems $\mathcal{X}, \mathcal{Y}$

The question

We observe two time series: $X = \{x_i\}, Y = \{y_i\}$ that is observation of two dynamic systems $\mathcal{X}, \mathcal{Y}$

X and Y are correlated

The question

We observe two time series: $X = \{x_i\}, Y = \{y_i\}$ that is observation of two dynamic systems $\mathcal{X}, \mathcal{Y}$

X and Y are correlated

$$\mathcal{X} \xrightarrow{?} \mathcal{Y}$$

$\mathcal{X}$ causes (drives) $\mathcal{Y}$-t?

The question

We observe two time series: $X = \{x_i\}, Y = \{y_i\}$ that is observation of two dynamic systems $\mathcal{X}, \mathcal{Y}$

X and Y are correlated

$$\mathcal{X} \xrightarrow{?} \mathcal{Y}$$

$\mathcal{X}$ causes (drives) $\mathcal{Y}$-t?

The question

or

$$\mathcal{X} \overset{?}{\leftarrow} \mathcal{Y}$$

$$\mathcal{X} \overset{?}{\leftrightarrow} \mathcal{Y}$$

The question

or

$$\mathcal{X} \overset{?}{\leftarrow} \mathcal{Y}$$

$$\mathcal{X} \overset{?}{\leftrightarrow} \mathcal{Y}$$

The question

Or is there a common driver $\mathcal{Z}$?

$$\begin{array}{ccc} & & \mathcal{X} \\ & \nearrow & \\ \mathcal{Z} & & \\ & \searrow & \\ & & \mathcal{Y} \end{array}$$

We need decision, but we can't do intervention, only observation, *Granger for SDS, Sugihara for DDS*

The question

Or is there a common driver $\mathcal{Z}$?

$$\begin{array}{ccc} & & \mathcal{X} \\ & \nearrow & \\ \mathcal{Z} & & \\ & \searrow & \\ & & \mathcal{Y} \end{array}$$

We need decision, but we can't do intervention, only observation, *Granger for SDS, Sugihara for DDS*

The unified theory

Fact

The unified theory handles deterministic and stochastic systems as well.

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \longrightarrow B$ a lamp flips to green, but not for $B \longrightarrow A$

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \rightarrow B$ a lamp flips to green, but not for $B \rightarrow A$ We find asymmetry.

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \rightarrow B$ a lamp flips to green, but not for $B \rightarrow A$ We find asymmetry.
3. We should show that green lamp implies if and only if $A \rightarrow B$.

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \rightarrow B$ a lamp flips to green, but not for $B \rightarrow A$ We find asymmetry.
3. We should show that green lamp implies if and only if $A \rightarrow B$.
4. We can separate cases: independence, uni- directional driving, mutual driving, there is a common driver (hidden common cause)

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \rightarrow B$ a lamp flips to green, but not for $B \rightarrow A$ We find asymmetry.
3. We should show that green lamp implies if and only if $A \rightarrow B$.
4. We can separate cases: independence, uni- directional driving, mutual driving, there is a common driver (hidden common cause)
5. We develop a method to quantify the probability that the lamp is green based on data.

Temporal causal discovery

1. How it works?
2. We find a law, a principle: If $A \rightarrow B$ a lamp flips to green, but not for $B \rightarrow A$ We find asymmetry.
3. We should show that green lamp implies if and only if $A \rightarrow B$.
4. We can separate cases: independence, uni- directional driving, mutual driving, there is a common driver (hidden common cause)
5. We develop a method to quantify the probability that the lamp is green based on data.

Temporal causal discovery

Stat method about green lamp

- detect,
- test,

Temporal causal discovery

Stat method about green lamp

- detect,
- test,
- separate cases:

Temporal causal discovery

Stat method about green lamp
- detect,
- test,
- separate cases:
  - visualize,

Temporal causal discovery

Stat method about green lamp
- detect,
- test,
- separate cases:
  - visualize,
  - test,

Temporal causal discovery

Stat method about green lamp
- detect,
- test,
- separate cases:
  - visualize,
  - test,
  - use Bayesian inference

Temporal causal discovery

Stat method about green lamp
- detect,
- test,
- separate cases:
  - visualize,
  - test,
  - use Bayesian inference
  - other tricky method...?

Temporal causal discovery

Stat method about green lamp
- detect,
- test,
- separate cases:
  - visualize,
  - test,
  - use Bayesian inference
  - other tricky method...?

related fork from us

Dr. Stippinger M., Dr. Zlatniczki Á, Dr. Benkő Zs., Dr. Stippinger M.,
Dr. Dr. Med. Fabó D., Dr. Dr. Med. Halász P., Dr. Dr. Med. Erőss L.,
Dr. Somogyvári Z, T.A.

The unified theory (Jakovác, T.)

But!

The atractor of a DS can be fractal, the dimension can be non integer, difficult to estimate.

The unified theory (Jakovác, T.)

But!
The atractor of a DS can be fractal, the dimension can be non integer, difficult to estimate.

The unified theory (Jakovác, T.)

But!

The atractor of a DS can be fractal, the dimension can be non integer, difficult to estimate.

The time delay embedding of a stochastic system fills up the space in any dimension.

The unified theory (Jakovác, T.)

But!
The atractor of a DS can be fractal, the dimension can be non integer, difficult to estimate.
The time delay embedding of a stochastic system fills up the space in any dimension.
The unified theory cures that problems.

The unified theory (Jakovác, T.)

But!

The atractor of a DS can be fractal, the dimension can be non integer, difficult to estimate.

The time delay embedding of a stochastic system fills up the space in any dimension.

The unified theory cures that problems.

Deterministic dynamic systems

$$\frac{d^{o_x}}{dt^{o_x}}x = f(x)$$
$$\frac{d^{o_y}}{dt^{o_y}}y = g(x, y)$$

e.g.

$$f(x) = a_0 + a_1 x + a_2 \frac{dx}{dt} + \ldots + a_{o_X} \frac{d^{o_X - 1}x}{dt^{o_X - 1}}.$$

Deterministic dynamic systems

$$\begin{aligned} \frac{d^{o_x}}{dt^{o_x}} x &= f(x) \\ \frac{d^{o_y}}{dt^{o_y}} y &= g(x, y) \end{aligned}$$

e.g.

$$f(x) = a_0 + a_1 x + a_2 \frac{dx}{dt} + \ldots + a_{o_X} \frac{d^{o_X - 1} x}{dt^{o_X - 1}}.$$

We assume all nice things.

Deterministic dynamic systems

$$\begin{aligned}\frac{d^{o_x}}{dt^{o_x}}x &= f(x)\\ \frac{d^{o_y}}{dt^{o_y}}y &= g(x,y)\end{aligned}$$

e.g.

$$f(x) = a_0 + a_1 x + a_2 \frac{dx}{dt} + \ldots + a_{o_X} \frac{d^{o_X - 1} x}{dt^{o_X - 1}}.$$

We assume all nice things.

Driving and observation

observation with τ time steps

$$
\begin{aligned}
x_{n+1} &= f(X^{n}_{n-o_X+1}) \\
y_{n+1} &= g(X^{n}_{n-o_X+1}, Y^{n}_{n-o_Y+1})
\end{aligned}
$$

$A_c = A_b^c = (A_c, A_{c-\tau}, A_{c-2\tau}...A_b)^T$ embedding.

Deterministic system

observation with τ time steps

$$\begin{aligned} x_{n+1} &= f(X^n_{n-o_X+1}) \\ y_{n+1} &= g(X^n_{n-o_X}, Y^n_{n-o_Y+1}) \end{aligned}$$

$A_c = A^c_b = (A_c, A_{c-\tau}, A_{c-2\tau} ... A_b)^T$ embedding.

Stochastic dynamic systems

$$
\begin{aligned}
x_{n+1} &= f(X_n, \xi_{n+1}) \\
y_{n+1} &= g(X_n, Y_n, \eta_{n+1})
\end{aligned}
$$

where ξ, η i.i.d. independent from the whole past $\left(X_0^{n-o_X-1}, Y_0^{n-o_Y-1}\right)$ and from each other.

Stochastic dynamic systems

$$
\begin{aligned}
x_{n+1} &= f(X_n, \xi_{n+1}) \\
y_{n+1} &= g(X_n, Y_n, \eta_{n+1})
\end{aligned}
$$

where ξ, η i.i.d. independent from the whole past $\left(X_0^{n-o_X-1}, Y_0^{n-o_Y-1}\right)$ and from each other.

When a DS is well defined (stoch.. or det.) ?

$$x_{n+1} = f(X_n, \xi_{n+1}) =: f_n(X_n)$$

$$\begin{aligned} x_{o_x+1} &= f\left(A_1^{o_X}, \xi_{o_x+1}\right) = f_1\left(A_{o_X}\right) \\ x_{n+1} &= f_n\left(X_n\right) = ... = f_n \circ f_{n-1} \circ \ldots \circ f_1(A_1^{o_x}) = F_n(A_{o_X}). \end{aligned}$$

X is well defined if f is known and the initial vector values are given

When a DS is well defined (stoch.. or det.) ?

$$x_{n+1} = f(X_n, \xi_{n+1}) =: f_n(X_n)$$

$$\begin{aligned} x_{o_x+1} &= f\left(A_1^{o_X}, \xi_{o_x+1}\right) = f_1\left(A_{o_X}\right) \\ x_{n+1} &= f_n\left(X_n\right) = \dots. = f_n \circ f_{n-1} \circ \dots \circ f_1(A_1^{o_x}) = F_n(A_{o_X}). \end{aligned}$$

X is well defined if f is known and the initial vector values are given $X_1^{o_X} = A_1^{o_X}$

When a DS is well defined (stoch.. or det.) ?

$$x_{n+1} = f(X_n, \xi_{n+1}) =: f_n(X_n)$$

$$\begin{aligned} x_{o_x+1} &= f\left(A_1^{o_X}, \xi_{o_x+1}\right) = f_1\left(A_{o_X}\right) \\ x_{n+1} &= f_n\left(X_n\right) = ... = f_n \circ f_{n-1} \circ ... \circ f_1(A_1^{o_x}) = F_n(A_{o_X}). \end{aligned}$$

X is well defined if f is known and the initial vector values are given $X_1^{o_X} = A_1^{o_X}$

$$df_X = o_X$$

When a DS is well defined (stoch.. or det.) ?

$$x_{n+1} = f(X_n, \xi_{n+1}) =: f_n(X_n)$$

$$\begin{aligned} x_{o_x+1} &= f\left(A_1^{o_X}, \xi_{o_x+1}\right) = f_1\left(A_{o_X}\right) \\ x_{n+1} &= f_n\left(X_n\right) = ... = f_n \circ f_{n-1} \circ ... \circ f_1(A_1^{o_x}) = F_n(A_{o_X}). \end{aligned}$$

X is well defined if f is known and the initial vector values are given $X_1^{o_X} = A_1^{o_X}$

$$df_X = o_X$$

And the driven?

We need $Y_1^{o_Y} = B_1^{o_Y}$ but $X_1^{o_X} = A_1^{o_X}$ as well.

$$\begin{aligned} y_{n+1} &= g_n(X_n, Y_n) \\ &= g_n \circ g_{n-1} \circ \ldots \circ g(F_n(A_{o_X}), B_{o_Y}) = G_n(F_n(A_{o_X}), B_{o_Y}) \end{aligned}$$

And the driven?

We need $Y_1^{o_Y} = B_1^{o_Y}$ but $X_1^{o_X} = A_1^{o_X}$ as well.

$$
\begin{aligned}
y_{n+1} &= g_n\left(X_n, Y_n\right) \\
&= g_n \circ g_{n-1} \circ \ldots \circ g(F_n(A_{o_X}), B_{o_Y}\ ) = G_n\left(F_n(A_{o_X}), B_{o_Y}\right)
\end{aligned}
$$

Lemma

The degrees if freedom of the driven is equal to the sum of ranks:

$$df_y = o_X + o_Y$$

And the driven?

We need $Y_1^{o_Y} = B_1^{o_Y}$ but $X_1^{o_X} = A_1^{o_X}$ as well.

$$\begin{aligned} y_{n+1} &= g_n\left(X_n, Y_n\right) \\ &= g_n \circ g_{n-1} \circ \ldots \circ g(F_n(A_{o_X}), B_{o_Y}\ ) = G_n\left(F_n(A_{o_X}), B_{o_Y}\right) \end{aligned}$$

Lemma

The degrees if freedom of the driven is equal to the sum of ranks:

$$df_y = o_X + o_Y$$

Task: let us determine the df of the system.

How can we test? Let it be DDS or SDS

$$X_{k-o_x}^{k-1} = (x_{k-1}, \dots . x_{k-o_x})^T \mapsto x_k$$

fix o_x number of values of $X_{k-o_x}^{k-1}$ that determines the distribution, the evolution of the system, (DSD and SDS as well) (picture as recursion). Values past to the fixed elements do not influence the evolution.

Task: let us determine the df of the system.

How can we test? Let it be DDS or SDS

$$X_{k-o_x}^{k-1} = (x_{k-1}, \ldots . x_{k-o_x})^T \mapsto x_k$$

fix o_x number of values of $X_{k-o_x}^{k-1}$ that determines the distribution, the evolution of the system, (DSD and SDS as well) (picture as recursion). Values past to the fixed elements do not influence the evolution. Markov property!

Task: let us determine the df of the system.

How can we test? Let it be DDS or SDS

$$X_{k-o_x}^{k-1} = (x_{k-1}, \dots x_{k-o_x})^T \mapsto x_k$$

fix o_x number of values of $X_{k-o_x}^{k-1}$ that determines the distribution, the evolution of the system, (DSD and SDS as well) (picture as recursion). Values past to the fixed elements do not influence the evolution. Markov property!

Even in the case of DDS! 1

Task: let us determine the df of the system.

How can we test? Let it be DDS or SDS

$$X_{k-o_x}^{k-1} = (x_{k-1}, \dots x_{k-o_x})^T \mapsto x_k$$

fix o_x number of values of $X_{k-o_x}^{k-1}$ that determines the distribution, the evolution of the system, (DSD and SDS as well) (picture as recursion). Values past to the fixed elements do not influence the evolution. Markov property! Even in the case of DDS! 1

Similarly for Y $df_Y = o_X + o_Y$ number of values has to be set.

Task: let us determine the df of the system.

How can we test? Let it be DDS or SDS

$$X_{k-o_x}^{k-1} = (x_{k-1}, \dots . x_{k-o_x})^T \mapsto x_k$$

fix o_x number of values of $X_{k-o_x}^{k-1}$ that determines the distribution, the evolution of the system, (DSD and SDS as well) (picture as recursion). Values past to the fixed elements do not influence the evolution. Markov property! Even in the case of DDS! 1

Similarly for Y $df_Y = o_X + o_Y$ number of values has to be set.

Summary (but not the end)

$$\max\{df_X, df_Y\} \leq df_{(X,Y)} \leq df_X + df_Y$$

Comparison

	WGC SDS	CCM DDS	DC DDAS	*df* both
$\nexists z, x \rightarrow y$	✓	✓	✓	✓ ✓
$\nexists z, x \leftarrow y$	✓	✓	✓	✓ ✓
$\nexists z, x \longleftrightarrow y$	✓	✓	✓	✓ ✓
$\exists z \left\{ x \nleftrightarrow y \right\}$	$\longleftrightarrow$	$\ldots$	✓	✓ ✓

Toy example

$$
\begin{aligned}
x_n &= \alpha Q_x x_{n-1} + g z_{n-1} + \beta \xi_{x,n} \\
y_n &= \alpha Q_y y_{n-1} + g z_{n-1} + \beta \xi_{y,n} \\
z_n &= \alpha Q_z z_{n-1} + \beta \xi_{z,n}
\end{aligned}
$$

$x, y, z \in \mathbb{R}^2$ we fix pairs!, $Q_.$ is a rotation matrix, $\alpha < 1$,
$df_x = df_y = 2 \times 2$, $df_z = 2$

Toy example

$$
\begin{aligned}
x_n &= \alpha Q_x x_{n-1} + g z_{n-1} + \beta \xi_{x,n} \\
y_n &= \alpha Q_y y_{n-1} + g z_{n-1} + \beta \xi_{y,n} \\
z_n &= \alpha Q_z z_{n-1} + \beta \xi_{z,n}
\end{aligned}
$$

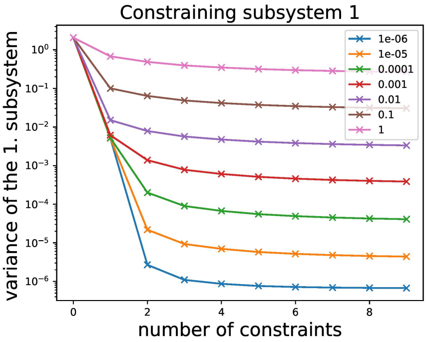
$x, y, z \in \mathbb{R}^2$ we fix pairs!, $Q_.$ is a rotation matrix, $\alpha < 1$, $df_x = df_y = 2 \times 2$, $df_z = 2$ z drives x and y and there is no driving between them.

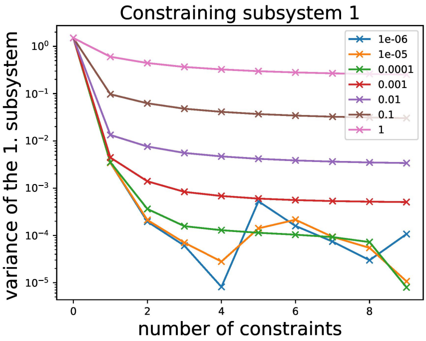
Toy example

$$\begin{aligned} x_n &= \alpha Q_x x_{n-1} + g z_{n-1} + \beta \xi_{x,n} \\ y_n &= \alpha Q_y y_{n-1} + g z_{n-1} + \beta \xi_{y,n} \\ z_n &= \alpha Q_z z_{n-1} + \beta \xi_{z,n} \end{aligned}$$

$x, y, z \in \mathbb{R}^2$ we fix pairs!, $Q_.$ is a rotation matrix, $\alpha < 1$, $df_x = df_y = 2 \times 2$, $df_z = 2$ z drives x and y and there is no driving between them.

Let us fix x

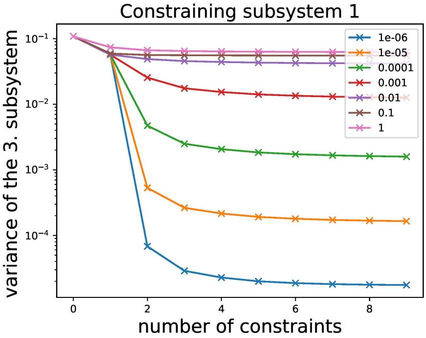




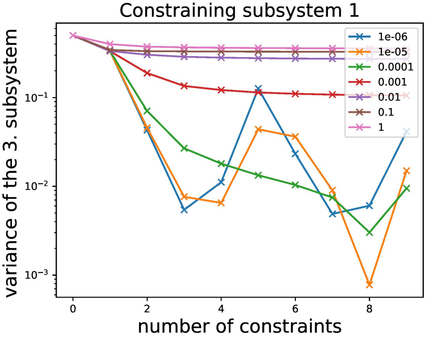
less stochastic more stochastic

remember, we set pairs of x!

We fix x and examine z



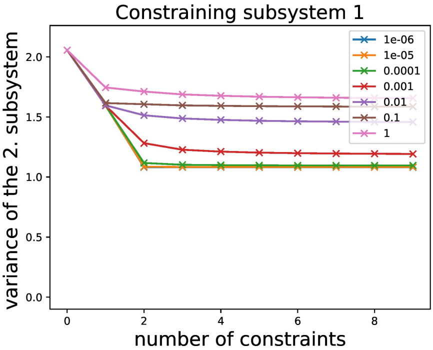
less stochastic



more stochastic

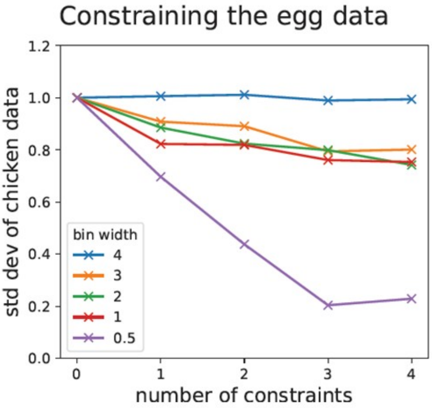
remember, we set pairs of x!

We fix x and examine y

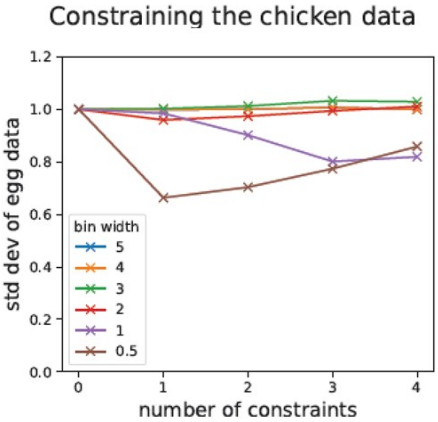


remember, we set pairs of x!

Chicken or egg?



Chicken or egg?



So we find assymmetry. $EGG \rightarrow Chicken$

Thanks
Q&A?