Functional causal models: Beyond linear instantaneous relations

Kun Zhang

Max-Planck Institute for Intelligent Systems
Tübingen, Germany
Causality vs. dependence: Examples

- Causality ➔ dependence ! dependence ➔ causality

X is a cause of Y iff
\[ \exists x_1 \neq x_2 \ P(Y | X \text{ set}=x_1) \neq P(Y | X \text{ set}=x_2) \]

X and Y are associated iff
\[ \exists x_1 \neq x_2 \ P(Y | X=x_1) \neq P(Y | X=x_2) \]
Causality vs. dependence: Examples

- Causality $\rightarrow$ dependence ! dependence $\rightarrow$ causality

\[ X \text{ is a cause of } Y \text{ iff } \exists x_1 \neq x_2 \ P(Y|X \text{ set}=x_1) \neq P(Y|X \text{ set}=x_2) \]

\[ X \text{ and } Y \text{ are associated iff } \exists x_1 \neq x_2 \ P(Y|X=x_1) \neq P(Y|X=x_2) \]
Causality vs. dependence: Examples

- Causality $\rightarrow$ dependence ! dependence $\rightarrow$ causality

X is a **cause** of Y iff
$$\exists x_1 \neq x_2 \; P(Y|X \text{ set}=x_1) \neq P(Y|X \text{ set}=x_2)$$

X and Y are **associated** iff
$$\exists x_1 \neq x_2 \; P(Y|X=x_1) \neq P(Y|X=x_2)$$
Causality vs. dependence: Examples

- Causality $\rightarrow$ dependence ! dependence $\rightarrow$ causality

$X$ is a *cause* of $Y$ iff
$\exists x_1 \neq x_2 \ P(Y | X \text{ set}=x_1) \neq P(Y | X \text{ set}=x_2)$

$X$ and $Y$ are *associated* iff
$\exists x_1 \neq x_2 \ P(Y | X=x_1) \neq P(Y | X=x_2)$
Brief history of causality:
Western philosophical tradition

- dates back at least to Aristotle
- Causality is not based on actual reasoning: only correlation can actually be perceived (David Hume, 1711-1776)
- One has to resort to **controlled experiments**
  - Manipulate a variable ‘ideally’ and see the response of the system
    \[ \Rightarrow \ldots \]
  - Usually impractical: smoking \(\rightarrow\) cough?
Brief history of causality: Eastern cultural tradition

- Illustrated Sutra of Cause and Effect (8th century)

- “coincidence” instead of causality (Carl Jung, 1920’s)
Potential applications

- Policy making in economics, climate analysis...
- Biology, brain connectivity analysis...
- Control, robust prediction / feature selection...
- For understanding learning problems, e.g., semi-supervised learning (Schölkopf et al., 2012)
- ...

$X \rightarrow Y \rightarrow Z$
Advances in the past decades: Computational causality

- In the past decades, under certain assumptions, it was made possible to derive causation from passively observed data (Pearl, Spirtes, Glymore, Scheines, Hoover et al.)
  - statistical data ⇒ causal structure
  - constraint-based approach
    - causal Markov assumption
    - faithfulness…

Contradicts classical claims ???
Outline

• Constraint-based causal discovery

• Functional causal model (mainly from 2005)
  • linear non-Gaussian causal model

• with necessary nonlinearities: Post-nonlinear causal model

X1 -- X2 -- X3

X1 → X2 → X3

X1 → X2 → X3
(even if very nonlinear)
From 1980’s...

• Constraint-based causal discovery
• Functional causal model (from 2005)
  • Linear, non-Gaussian causal model
• Post-nonlinear causal model
Causal structure vs. statistical independence
(Spiritse, Pearl, et al.)

causal structure
(causal graph)
\[ Y \rightarrow X \rightarrow Z \]
\[ Y \perp\!\!\!\!\!\!\perp X \perp\!\!\!\!\!\!\perp Z \]

Recall:
\[ Y \perp\!\!\!\!\!\!\perp Z \Leftrightarrow P(Y|Z) = P(Y); Y \perp\!\!\!\!\!\!\perp Z|X \Leftrightarrow P(Y|Z,X) = P(Y|X) \]
Causal structure vs. statistical independence
(Spirtes, Pearl, et al.)

Causal Markov condition: each variable is independent of its non-descendants (non-effects) conditional on its parents (direct causes).

causal structure (causal graph)
\[ Y \rightarrow X \rightarrow Z \]

\[ Y \perp\!\!\!\!\!\!\!\perp Z \iff P(Y|Z)=P(Y); \quad Y \perp\!\!\!\!\!\!\!\perp Z|X \iff P(Y|Z,X)=P(Y|X) \]

Recall:

Thursday, 13 June 2013
Causal structure vs. statistical independence
(Spirit, Pearl, et al.)

Causal Markov condition: each variable is ind. of its non-descendants (non-effects) conditional on its parents (direct causes)

causal structure
(causal graph)

\[ Y \rightarrow X \rightarrow Z \]

\[ Y \perp\!\!\!\!\!\perp Z \iff P(Y|Z) = P(Y); Y \perp\!\!\!\!\!\perp Z|X \iff P(Y|Z,X) = P(Y|X) \]

Thursday, 13 June 2013
Causal structure vs. statistical independence (Spirtes, Pearl, et al.)

Causal Markov condition: each variable is independent of its non-descendants (non-effects) conditional on its parents (direct causes).

Causal structure (causal graph)
\[ Y \rightarrow X \rightarrow Z \]
\[ Y \perp\!\!\!\!\!\!\perp X \perp\!\!\!\!\!\!\perp Z \]

Statistical independence(s)
\[ Y \quad Z \mid X \]

Recall:
\[ Y \perp\!\!\!\!\!\!\perp Z \iff P(Y|Z) = P(Y); Y \perp\!\!\!\!\!\!\perp Z|X \iff P(Y|Z,X) = P(Y|X) \]
Causal structure vs. statistical independence
(Spirtes, Pearl, et al.)

Causal Markov condition: each variable is ind.
of its non-descendants (non-effects) conditional on 
its parents (direct causes)

causal structure
(causal graph)

\[ \begin{align*} Y \rightarrow X \rightarrow Z \\
Y \perp\!
\perp X \perp\!
\perp Z \end{align*} \]

Statistical independence(s)

\[ Y \perp\!
\perp Z \mid X \]

Recall:
\[ Y \perp\!
\perp Z \Leftrightarrow P(Y|Z) = P(Y); Y \perp\!
\perp Z|X \Leftrightarrow P(Y|Z,X) = P(Y|X) \]
Causal structure vs. statistical independence
(Spirit, Pearl, et al.)

Causal Markov condition: each variable is ind. of its non-descendants (non-effects) conditional on its parents (direct causes)

causal structure (causal graph)

$Y \rightarrow X \rightarrow Z$

$Y \perp\!\!\!\!\!\!\!\perp X -- Z$

Statistical independence(s)

$Y \rightleftharpoons Z | X$

Faithfulness: all observed (conditional) independencies are entailed by Markov condition in the causal graph

Recall: $Y \rightleftharpoons Z \iff P(Y|Z) = P(Y)$; $Y \rightleftharpoons Z|X \iff P(Y|Z,X) = P(Y|X)$
Causal structure vs. statistical independence
(Spiritse, Pearl, et al.)

Causal Markov condition: each variable is independent of its non-descendants (non-effects) conditional on its parents (direct causes).

Causal structure (causal graph)
\[ Y \rightarrow X \rightarrow Z \]

Y -- X -- Z ?

Statistical independence(s)
\[ Y \perp\!\!\!\!\!\!\!\!\perp Z | X \]

Faithfulness: all observed (conditional) independencies are entailed by Markov condition in the causal graph.

Recall:
\[ Y \perp\!\!\!\!\!\!\!\!\perp Z \iff P(Y|Z) = P(Y); Y \perp\!\!\!\!\!\!\!\!\perp Z|X \iff P(Y|Z,X) = P(Y|X) \]

Thursday, 13 June 2013
Constraint-based causal discovery

• Theorem: if \((G, P)\) satisfies faithfulness, then there is an edge between \(X\) and \(Y\) iff \(X \perp Y\) given any set of variables

• uses (conditional) independence constraints to find the candidate causal structures

• example: PC algorithm (Spirtes & Glymour, 1991)
Markov equivalence class

pattern Y -- X -- Z

same adjacencies

→ if all agree on orientation; -- if disagree

might be unique: v-structure
Constraint-based method: An inverse problem

- \{\text{local causal structures}\} \rightarrow \{\text{conditional independences}\}

\[
\begin{array}{ccc}
\text{X} & \text{Y} & \text{Z} \\
\text{Z} & \text{X} \rightarrow & \text{Y} \\
\text{X} & \text{Z} \rightarrow & \text{Y} \\
\text{X} & \text{Y} & \text{Z} \\
\text{X} & \text{Y} & \text{Z} \\
\end{array}
\]

\[
\begin{array}{c}
\emptyset \\
\text{X \perp Y} \\
\text{X \perp Z | Y} \\
\end{array}
\]
Constraint-based method: An inverse problem

- \{\text{local causal structures}\} \rightarrow \{\text{conditional independences}\}

\[
\begin{array}{ccc}
X & Y & Z \\
X & Z & Y \\
X & Y & Z \\
X & Y & Z \\
\end{array}
\]

faithfulness

\[
\begin{array}{c}
\emptyset \\
X \perp Y \\
X \perp Z | Y \\
\end{array}
\]
Constraint-based method: An inverse problem

- \{\text{local causal structures}\} \rightarrow \{\text{conditional independences}\}
Constraint-based method: An inverse problem

- \{local causal structures\} \rightarrow \{conditional independences\}

\begin{table}
\begin{tabular}{c|c|c}
\hline
X & Y & Z \\
\hline
X & Z & Y \\
X & Z & Y \\
X & Z & Y \\
X & Z & Y \\
\hline
\end{tabular}
\end{table}

\textit{X \perp Y}

\textit{X \perp Z | Y}

\textit{faithfulness}

\textit{two-variable case?}

\textit{equivalence class}
Constraint-based method: An inverse problem

- \{\text{local causal structures}\} \rightarrow \{\text{conditional independences}\}

- Instead, we try to directly identify local causal structures with functional causal models

- Instead, we try to directly identify local causal structures with functional causal models

- Two-variable case?

- Equivalence class

- Faithfulness
Outline

• Constraint-based causal discovery

• Functional causal model (from 2005)
  • Linear, non-Gaussian causal model
  • Post-nonlinear causal model

• \( X_1 \rightarrow X_2 \rightarrow X_3 \) (if linear)

• \( X_1 \rightarrow X_2 \rightarrow X_3 \) (even if very nonlinear)
Functional causal model (Pearl et al.)

- generative function model for continuous variables
  \[ x_i = f_i(pa_i, e_i), \quad i = 1, ..., n \]

- in econometrics, social sciences...

- well-defined examples

- Granger causality: effects follow causes in a linear form

- LiNGAM: linear, non-Gaussian and acyclic causal model (Shimizu et al., 2006)
FCM: A general view

- Without constraints on $f$, for given $(X, Y)$, both $y = f_1(x, e)$ with $E_{\|X}$ and $x = f_2(y, e_1)$ with $E_{\|Y}$ are possible.

- with a Gram-Schmidt-orthogonalization procedure (Darmois, 1951)

  \[ \tilde{x} = \text{cdf}(x_1), \text{ so } \tilde{x} \sim U(0,1); \]
  \[ e = \text{cdf}(y \mid \tilde{x}) = \int_{-\infty}^{x_2} p_{\tilde{x},y}(\tilde{x},t)dt. \]

  Then $(x, y) \Rightarrow (\tilde{x}, e)$, with $E_{\|X}$. 

Thursday, 13 June 2013
Suppose we observe the data
A universal way to construct “trivial” FCMs

- $e' = h \circ \text{CCDF}_{Y|X}(y|x)$ always independent from $X$
- Functional causal model: $y = \text{CCDF}_{Y|X}^{-1} \circ h^{-1}(e')$ for any $x$
- how to make it identifiable (break the symmetry)?
General FCMs: independence vs. likelihood

- relating mutual information $I$ and likelihood $l$:
  
  $$l_{X \rightarrow Y}(\beta) = \sum_{i=1}^{n} \log P_f(x_i, y_i) = \sum_{i=1}^{n} \log P(X = x_i, Y = y_i) - I(X, E; \beta)$$

- If $X \rightarrow Y$ follows the model:
  
  $$l_{X \rightarrow Y}(\beta^*) - l_{Y \rightarrow X}(\beta_Y^*) = I(Y, E_Y; \beta_Y^*)$$

- also hold for more than two variables
A basic functional causal model

• Constraint-based causal discovery

• Functional causal models (from 2005)
  • Linear, non-Gaussian acyclic causal model
  • Post-nonlinear causal model
LiNGAM model

- **linear, non-Gaussian, acyclic causal model** (LiNGAM) (Shimizu et al., 2006):

\[
x_i = \sum_{j: \text{parents of } i} b_{ij} x_j + e_i \quad \text{or} \quad x = Bx + e
\]

- disturbances (errors) \(e_i\) are non-Gaussian (or at most one is Gaussian) and mutually indep.

- example:

\[
\begin{align*}
x_2 &= e_2, \\
x_3 &= 0.5x_2 + e_3, \\
x_1 &= -0.2x_2 + 0.3x_3 + e_1.
\end{align*}
\]
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \quad \text{and} \quad y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank

Thursday, 13 June 2013
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]

\[ y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]
\[ y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]
\[ y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank

ICA system
observed signals
output: as independent as possible
demixing
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]
\[ y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]

\[ y = W \cdot x \]

- assumptions in ICA
  - at most one of \( s_i \) is Gaussian
  - \( m \geq n \), and \( A \) is of full column rank

\( \text{ICA system} \)

\( \text{output: as independent as possible} \)

\( \text{de-mixing} \)

\( \text{observed signals} \)
ICA: A well-known technique making use of non-Gaussianity

\[ x = A \cdot s \]

unknown mixing system

\[ y = W \cdot x \]

ICA system

• assumptions in ICA
  • at most one of \( s_i \) is Gaussian
  • \( m \geq n \), and \( A \) is of full column rank

\[ y_1, \ldots, y_n \]

output: as independent as possible

\[ x_1, \ldots, x_m \]

observed signals

\[ s_1, \ldots, s_n \]

independent sources
LiNGAM analysis by ICA

- LiNGAM: \( x = Bx + e \Rightarrow e = (I - B)x \)
- \( B \) has a special structure: **acyclic relations**
- ICA: \( y = Wx \)
- \( B \) can then be seen from \( W \) by permutation and re-scaling
- e.g.

\[
\begin{bmatrix}
  y_1 \\
  y_3 \\
  y_2 
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & 0 \\
  -0.5 & 1 & 0 \\
  0.2 & -0.3 & 1 
\end{bmatrix}
\begin{bmatrix}
  x_2 \\
  x_3 \\
  x_1 
\end{bmatrix}
\]

\[\Leftrightarrow \begin{cases} 
    x_2 = y_1 \\
    x_3 = 0.5x_2 + y_3 \\
    x_1 = -0.2x_2 + 0.3x_3 + y_2 
\end{cases} \]

So we have the causal relation:
Related work & applications

- ICA with sparse connections (Zhang et al., 2008); Direct LiNGAM (Shimizu et al., 2009)

- with mild nonlinear distortion allowed; application in finance (Zhang & Chan, 2006 & 2008)

- extended Granger causality analysis for time series (Hyvärinen et al., 2010; Zhang and Hyvärinen, 2009)
Now comes...

- Constraint-based causal discovery
- Functional causal model (from 2005)
  - Linear, non-Gaussian acyclic causal model
- Post-nonlinear (PNL) causal model
Three Effects usually encountered in a causal model (Zhang & Hyvärinen, 2009)

• Without prior knowledge, the assumed model is expected to be
  • general enough: adapted to approximate the true generating process
  • identifiable: asymmetry in causes and effects

• represented by post-nonlinear causal model with inner additive noise
PNL causal model with inner additive noise

- acyclic data-generating process

\[ x_i = f_{i,2}(f_{i,1}(pa_i) + e_i) \]

- two-variable case

- \( x_1 \rightarrow x_2: x_2 = f_{2,2}(f_{2,1}(x_1) + e_2) \)

- \( pa_i \): parents (causes) of \( x_i \)

- \( f_{i,2} \): assumed to be continuous and invertible

- \( f_{i,1} \): not necessarily invertible

- \( e_i \): noise/disturbance: independent from \( pa_i \)
Special cases of PNL causal model

\[ x_i = f_{i,2} (f_{i,1} (pa_i) + e_i) \]

- If \( f_{i,1} \) and \( f_{i,2} \) are both linear
- At most one of \( e_i \) is Gaussian: LiNGAM (Shimizu et al., 2006)
- All of \( e_i \) are Gaussian: linear Gaussian case (Spirtes, Pearl et al.)
- If \( f_{i,2} \) is identity: nonlinear causal discovery with additive noise models (Hoyer et al., 2009, Zhang 2009b)
Identifiability in two-variable case

- Is the causal direction implied by the model unique?
- We tackle this problem by a proof of contradiction
  - Assume both $x_1 \rightarrow x_2$ and $x_1 \leftarrow x_2$ satisfy PNL model
  - One can then find all non-identifiable cases
Identifiability: A mathematical result

Theorem 1

- Assume
  \[x_2 = f_2(f_1(x_1) + e_2),\]
  \[x_1 = g_2(g_1(x_2) + e_1),\]

- Further suppose that involved densities and nonlinear functions are third-order differentiable, and that \(p_{e_2}\) is unbounded,

- For every point satisfying \(\eta_2'' h' \neq 0\), we have
  \[
  \eta_1''' - \frac{\eta_1'' h''}{h'} = \left(\frac{\eta_2'''}{\eta_2''} - 2\eta_2''\right) \cdot h'' - \frac{\eta_2''}{\eta_2''} \cdot h' \eta_1'' + \eta_2' \cdot \left(\frac{h'''}{h'} - \frac{h''}{h'}^2\right).
  \]

- Obtained by using the fact that the Hessian of the logarithm of the joint density of independent variables is diagonal everywhere (Lin, 1998)

- It is not obvious if this theorem holds in practice…
Can we find conditions easy to verify?

Fortunately, the DE can be re-written as a bilinear functional equation

$$\Phi_1(t_1)\Psi_1(e_2) + \Phi_2(t_1)\Psi_2(e_2) + \Phi_3(t_1)\Psi_3(e_2) + \Phi_4(t_1)\Psi_4(e_2) = 0,$$

where

$$\Phi_1(t_1) = \eta_1''' - \frac{\eta_1'''h''}{h'}, \quad \Phi_2(t_1) = h''' - \frac{h''^2}{h'}, \quad \Phi_3(t_1) = h'h'', \quad \Phi_4(t_1) = h'\eta_1''',$$

$$\Psi_1(e_2) = -1, \quad \Psi_2(e_2) = \eta'_2, \quad \Psi_3(e_2) = \frac{\eta_2\eta'''_2}{\eta''_2} - 2\eta''_2, \quad \Psi_4(e_2) = -\frac{\eta''''_2}{\eta''_2}.$$

One can find the solutions in closed form or analyze their behaviour in all possible cases.

- **Simple cases:** $\Phi_3(t_1) \equiv 0$, $\Psi_4(e_2) \equiv 0$, or $\Phi_2(t_1) \equiv 0$.
- **Complex cases:**
  - **Solution 1:**
    $$\Phi_1 = A_1\Phi_3 + A_2\Phi_4, \quad \Phi_2 = A_3\Phi_3 + A_4\Phi_4,$$
    $$\Psi_3 = -A_1\Psi_1 - A_3\Psi_2, \quad \Psi_4 = -A_2\Psi_1 - A_4\Psi_2,$$
  - **Solution 2:**
    $$\Phi_1 = B_1\Phi_3, \quad \Phi_2 = B_2\Phi_3, \quad \Phi_4 = B_3\Phi_3,$$
    $$\Psi_3 = -B_1\Psi_1 - B_2\Psi_2 - B_3\Psi_4,$$
  - **Solution 3:**
    $$\Psi_2 = C_1\Psi_1, \quad \Psi_3 = C_2\Psi_1, \quad \Psi_4 = C_3\Psi_1,$$
    $$\Phi_1 = -C_1\Phi_2 - C_2\Phi_3 - C_3\Phi_4,$$

where $A_i$, $B_i$, and $C_i$ are arbitrary constants.
Finally: All non-identifiable cases

Log-mixed-linear-and-exponential:
\[ \log p_v = c_1 e^{c_2 v} + c_3 v + c_4 \]

\[ (\log p_v)' \rightarrow c_1 (c_1 \neq 0), \]
\[ \text{as } v \rightarrow -\infty \text{ or as } v \rightarrow +\infty \]

Table 1: All situations in which the PNL causal model is not identifiable.

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_c )</td>
<td>( p_t )</td>
<td>( h )</td>
<td>Remark</td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>Gaussian</td>
<td>linear</td>
<td>( h_1 ) also linear</td>
<td></td>
</tr>
<tr>
<td>log-mix-lin-exp</td>
<td>log-mix-lin-exp</td>
<td>linear</td>
<td>( h_1 ) strictly monotonic, and ( h'_1 \rightarrow 0 ), as ( z_2 \rightarrow +\infty ) or as ( z_2 \rightarrow -\infty )</td>
<td></td>
</tr>
<tr>
<td>log-mix-lin-exp</td>
<td>one-sided asymptotically exponential (but not log-mix-lin-exp)</td>
<td>( h ) strictly monotonic, and ( h' \rightarrow 0 ), as ( t_1 \rightarrow +\infty ) or as ( t_1 \rightarrow -\infty )</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>log-mix-lin-exp</td>
<td>generalized mixture of two exponentials</td>
<td>Same as above</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>generalized mixture of two exponentials</td>
<td>two-sided asymptotically exponential</td>
<td>Same as above</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

\[ p_v \propto (c_1 e^{c_2 v} + c_3 e^{c_4 v})^{c_5} \]
Corollaries easy to verify

• Corollary 1: If $p_{e2}$ is not Gaussian, nor log-mix-lin-exp, nor a generalized mixture of two exponentials, then the PNL causal model is identifiable

\[ x_2 = f_2(f_1(x_1) + e_2) \]

• Corollary 2: If function $f_1$ is not invertible, then the PNL causal model is identifiable
Method for distinguishing cause from effect

- Examine if $x_1 \rightarrow x_2$ holds
- Examine if $x_2 \rightarrow x_1$ holds
- Draw conclusions
  - Only one of them holds 😊😊
  - Both hold: they could not be distinguished by PNL
    - Additional information of the nonlinearities, such that the smoothness, nonlinear distortion level, etc. may be helpful
  - If neither of them holds, data do not follow PNL, or confounders have significant effects 😞😞
Method to examine if $x_1 \rightarrow x_2$

- If $x_1 \rightarrow x_2$, i.e., $x_2 = f_{2,2}(f_{2,1}(x_1) + e_2)$, we have $e_2 = f_{2,2}^{-1}(x_2) - f_{2,1}(x_1)$ is ind. from $x_1$

- Two-step procedure to examine if $x_1 \rightarrow x_2$
  - Step 1: makes $y_2 = g_2(x_2) - g_1(x_1)$ and $x_1$ as ind. as possible, such that $y_2$ provides $\hat{e}_2$

- Step 2: uses independence tests (Gretton, et al., 2008) to verify if $x_1$ and $\hat{e}_2$ are ind.
Application on real data

• applied on “CausalEffectPairs”
  • 80 data sets for cause-effect pairs; each contains realizations of two variables
  • Causal direction is obvious to non-experts, but background information is hidden for participants
  • Goal: to distinguish cause from effect of the two variables
Performance

- with MLP and automatic initialization

- Local optima due to MLP’s. Performance improved with specific preprocessing for each pair

(figure adapted from Janzing et al., 2012)
By warped Gaussian Processes

• Aim at estimating $f$ and $e$ in $x_2 = f_{2,2}(f_{2,1}(x_1) + e_2)$

• Using a Gaussian process prior for $f_{2,1}$

• $\Rightarrow$ warped Gaussian processes (Snelson et al., 2004)

• We also consider non-Gaussian noise (modeled by the mixture of Gaussians)

• Performance on cause-effect pairs
  • with Gaussian noise: $52/(52+19) \sim 73\%$
  • with MoG noise: $56/(56+18) \sim 76\%$

• MATLAB code available upon request
Data Set 1

(a) $y_1$ vs $y_2$ under hypothesis $x_1 \rightarrow x_2$

(b) $y_1$ vs $y_2$ under hypothesis $x_2 \rightarrow x_1$

$x_1$ vs. its nonlinear effect on $x_2$

$x_2$ vs. $f_{2,2}^{-1}(x_2)$

Independence test results on $y_1$ and $y_2$ with different assumed causal relations

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$x_1 \rightarrow x_2$ assumed</th>
<th>$x_2 \rightarrow x_1$ assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 0.01$</td>
<td>$\alpha = 0.01$</td>
</tr>
<tr>
<td>#1</td>
<td>$2.3 \times 10^{-3}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>$1.7 \times 10^{-3}$</td>
<td>$6.5 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Independent test results on $y_1$ and $y_2$ with different assumed causal relations.
Data Set 8

(a) $y_1$ vs $y_2$ under hypothesis $x_1 \rightarrow x_2$

(b) $y_1$ vs $y_2$ under hypothesis $x_2 \rightarrow x_1$

$y_2$ (estimate of $e_2$)

$y_2$ (estimate of $e_1$)

$x_2$ vs. $f^{-1}_{2,2}(x_2)$

Nonlinear effect of $x_1$

$x_1$ vs. its nonlinear effect on $x_2$
Summary: Post-nonlinear causal model

• very general + identifiable
  • Both the nonlinear effect of the cause and the sensor distortion usually exist

• clear physical interpretations of the data generating process (causal influence)
Constraint-based vs. functional causal model based causal discovery

- **Constraint-based approaches**
  - + might avoid assuming the form of $f$ provided powerful CI tests
  - - info loss (underdetermination, orientation error propagation)

- **Functional causal model based approaches**
  - + could directly determine local causal structures (identifiable), & is interpretable and facilitates prediction
  - - how to find the form or appropriate knowledge of $f$?!
  - + both could be generalized to the confounder case
  - Usually both involve multiple testing (possible for functional causal models to avoid, using likelihood as the score)
Thank you!

Thanks also go to

- Bernhard Schölkopf
- Laiwan Chan
- Lei Xu
- Patrik Hoyer
- Joris Mooij
- Jakob Zscheischler

- Aapo Hyvärinen
- Dominik Janzing
- Kenji Fukumizu
- Jonas Peters
- Shohei Shimizu
- Eleni Sgouritsa

Thursday, 13 June 2013
Re-consider the examples...

- Causality $\rightarrow$ dependence! dependence $\rightarrow$ causality
Re-consider the examples...

- Causality $\rightarrow$ dependence! dependence $\rightarrow$ causality
Re-consider the examples...

- Causality $\rightarrow$ dependence! dependence $\rightarrow$ causality
Re-consider the examples...

- Causality $\rightarrow$ dependence! dependence $\rightarrow$ causality

Time table

...
Some References

- K. Zhang & L. Chan. Extensions of ICA for causal discovery in Hong Kong stock market, ICONIP 2006
- K. Zhang & A. Hyvärinen. Causal discovery with nonlinear acyclic causal models, JMLR W&CP, 2010