Causal models for decision making via integrative inference

Philipp Geiger

MPI for Intelligent Systems and University of Stuttgart

20th June 2017

Causal models for decision making via integrative inference – outline

Background in causal models

Motivation and overview over thesis (3 projects)

Project 1

Project 2

Conclusions

Advertisement decision example

Setup: Ad department of some web shop

- $A \in \{0, 1\}$: send letter to some person
- $B \in \{0, 1\}$: the person buys something

Advertisement decision example

Setup: Ad department of some web shop

- $A \in \{0, 1\}$: send letter to some person
- $B \in \{0, 1\}$: the person buys something

Question:

Distribution of B after sending/not sending ad letter in current situation? (Have some goal w.r.t. A, B.)

Equivalent to: causal influence of A on B

Denote it by $P(B|\mathrm{do}\, A = a), a = 0, 1$

Advertisement decision example – inference

Asm: In past situation, we recorded $P(A, B, C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

Advertisement decision example – inference

Asm: In past situation, we recorded $P(A, B, C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

Asm: $P(A, B, C)$ generated by these *structural equations* ("mechanisms"):

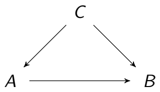
$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$

Advertisement decision example – inference

Asm: In past situation, we recorded $P(A, B, C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

Asm: $P(A, B, C)$ generated by these *structural equations* (“mechanisms”):

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$

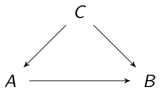


Advertisement decision example – inference

Asm: In past situation, we recorded $P(A, B, C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

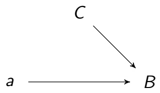
Asm: $P(A, B, C)$ generated by these *structural equations* ("mechanisms"):

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$



Def: the effect $P(B|\mathrm{do}\, A{=}a)$ is given by distribution of B after *intervening* on A, setting to constant a:

$$B := f_B(a, C, U_B)$$
$$C := f_C(U_C)$$



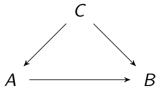
(Asm: other mechanisms invariant)

Advertisement decision example – inference

Asm: In past situation, we recorded $P(A, B, C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

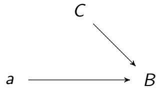
Asm: $P(A, B, C)$ generated by these *structural equations* ("mechanisms"):

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$



Def: the effect $P(B|\mathrm{do}\, A{=}a)$ is given by distribution of B after *intervening* on A, setting to constant a:

$$B := f_B(a, C, U_B)$$
$$C := f_C(U_C)$$



(Asm: other mechanisms invariant)

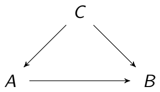
Conclude causal effect: $P(B{=}b|\mathrm{do}\, A{=}a) = \sum_c P(b|a, c)P(c)$

Advertisement decision example – inference

Asm: In past situation, we recorded $P(A,B,C)$, where C denotes, say, gender (recall: A: send ad letter; B: buying)

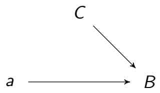
Asm: $P(A,B,C)$ generated by these *structural equations* ("mechanisms"):

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$



Def: the effect $P(B|\mathrm{do}\, A{=}a)$ is given by distribution of B after *intervening* on A, setting to constant a:

$$B := f_B(a, C, U_B)$$
$$C := f_C(U_C)$$



(Asm: other mechanisms invariant)

Conclude causal effect: $P(B{=}b|\mathrm{do}\, A{=}a) = \sum_c P(b|a,c)P(c)$

(drop $A := f_A(C,U_A)$ $\Leftrightarrow$ drop $P(a|c)$ from $P(c)P(a|c)P(b|a,c)$)

General definition, hidden confounding

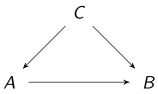
$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$

Def. [Pearl 2000]: Given a set of variables V a *causal model* M consists of, for each $X \in V$:

- a *structural equation* $X := f_X(PA_X, U_X)$ for $PA_X \subset V$
- a *background variable* U_X, with distribution $P(U_X)$

General definition, hidden confounding

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$



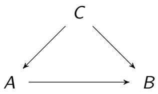
Def. [Pearl 2000]: Given a set of variables V a *causal model* M consists of, for each $X \in V$:

- a *structural equation* $X := f_X(PA_X, U_X)$ for $PA_X \subset V$
- a *background variable* U_X, with distribution $P(U_X)$

Causal model induces: DAG and *post-interventional* $P(Y|\mathrm{do}\, X{=}x)$

General definition, hidden confounding

$$A := f_A(C, U_A)$$
$$B := f_B(A, C, U_B)$$
$$C := f_C(U_C)$$



Def. [Pearl 2000]: Given a set of variables V a *causal model* M consists of, for each $X \in V$:

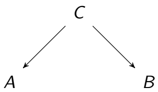
- a *structural equation* $X := f_X(PA_X, U_X)$ for $PA_X \subset V$
- a *background variable* U_X, with distribution $P(U_X)$

Causal model induces: DAG and *post-interventional* $P(Y|\mathrm{do}\, X{=}x)$

Problem of hidden confounding: C hidden $\Rightarrow$ $P(B|\mathrm{do}\, a)$ undeterm.

General definition, hidden confounding

$A := C$
$B := C$
$C := f_C(U_C)$



Def. [Pearl 2000]: Given a set of variables V a *causal model M* consists of, for each $X \in V$:

- a *structural equation* $X := f_X(PA_X, U_X)$ for $PA_X \subset V$
- a *background variable* U_X, with distribution $P(U_X)$

Causal model induces: DAG and *post-interventional* $P(Y|\mathrm{do}\, X{=}x)$

Problem of hidden confounding: C hidden $\Rightarrow P(B|\mathrm{do}\, a)$ undeterm.

In particular: $P(b|\mathrm{do}\, a) \neq P(b|a)$
E.g.: (model above) $\Rightarrow P(b|\mathrm{do}\, a) = P(b) \neq \delta_a(b) = P(b|a)$

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

→ Thesis: causal models as a basis for *data-driven decision making*

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

→ Thesis: causal models as a basis for *data-driven decision making*

Prev. causal models research: focus on exact and observational/iid

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

→ Thesis: causal models as a basis for *data-driven decision making*

Prev. causal models research: focus on exact and observational/iid

→ Thesis: *integrative* and *approximative* inference of causal models

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

$\rightarrow$ Thesis: causal models as a basis for *data-driven decision making*

Prev. causal models research: focus on exact and observational/iid

$\rightarrow$ Thesis: *integrative* and *approximative* inference of causal models

Proj.	integrates (besides obs.)	for decisions in	publ. at
1	• temporal knowledge		ICML'15
2	• bounds on confounding	• advertisement	UAI '14
3 $\notin$ talk	• partial models • system specifications • ...	• debugging and • bidding in cloud computing	arXiv

Motivation and overview over thesis (3 projects)

Causal questions relevant for decisions in medicine, business, etc.

→ Thesis: causal models as a basis for *data-driven decision making*

Prev. causal models research: focus on exact and observational/iid

→ Thesis: *integrative* and *approximative* inference of causal models

Proj.	integrates (besides obs.)	for decisions in	publ. at
1	• temporal knowledge		ICML'15
2	• bounds on confounding	• advertisement	UAI '14
3 $\notin$ talk	• partial models • system specifications • ...	• debugging and • bidding in cloud computing	arXiv

Contributions: in the form of theorems, algorithms, and conceptual

Causal models for decision making via integrative inference – outline

Background in causal models

Motivation and overview over thesis

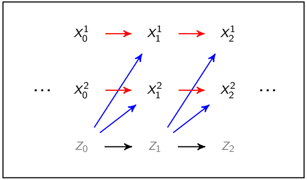
Project 1: Causal inference from confounded time series

Project 2: Causal inference in i.i.d. settings by bounding confounding

Conclusions

Overview: problem and contributions

Goal: causal model of dynamical system



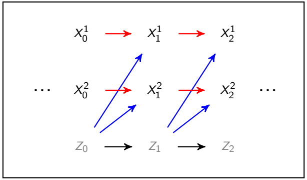
Given: time series $X_{0:L}$

Background: time → causal ordering ☺
[Granger 1969].
But hidden $Z_{0:L}$ ☹ – barely studied

Example: X_0^1: cheese price at $t = 0$; X_1^2: butter price at $t = 1$

Overview: problem and contributions

Goal: causal model of dynamical system



Given: time series $X_{0:L}$

Background: time $\rightarrow$ causal ordering ☺ [Granger 1969].
But hidden $Z_{0:L}$ ☹ – barely studied

Example: X_0^1: cheese price at $t = 0$; X_1^2: butter price at $t = 1$

Contributions:

- Three theorems: conditions for identifiability of influences in spite of hidden confounders in VAR processes (approximately)
- Two propositions: genericity of several conditions
- Two algorithms: estimation from finite data (under cond.)

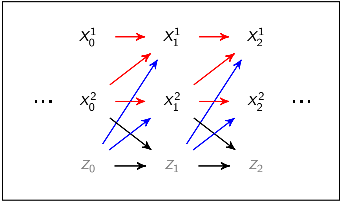
Spotlight: non-Gaussian identifiability – theorem

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Structural equations

If

- $\begin{pmatrix} X_t \\ Z_t \end{pmatrix}$ is VAR process



Example causal DAG

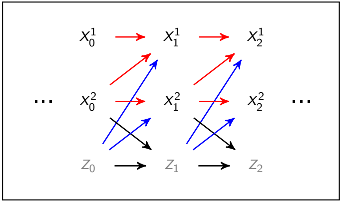
Spotlight: non-Gaussian identifiability – theorem

If

- $\begin{pmatrix} X_t \\ Z_t \end{pmatrix}$ is VAR process
- N_t are independent and non-Gaussian
- Plus *generic* further assumptions.

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Structural equations

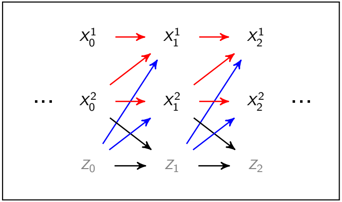


Example causal DAG

Spotlight: non-Gaussian identifiability – theorem

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Structural equations



Example causal DAG

If

- $\begin{pmatrix} X_t \\ Z_t \end{pmatrix}$ is VAR process
- N_t are independent and non-Gaussian
- Plus *generic* further assumptions.

Then

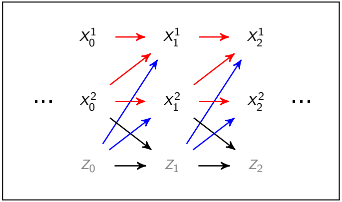
given only $P\left((X_t)_{t\in\mathbb{Z}}\right)$,

the matrix B is *uniquely identifiable*

Spotlight: non-Gaussian identifiability – theorem

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Structural equations



Example causal DAG

If

- $\begin{pmatrix} X_t \\ Z_t \end{pmatrix}$ is VAR process
- N_t are independent and non-Gaussian
- Plus *generic* further assumptions.

Then

given only $P\left((X_t)_{t \in \mathbb{Z}}\right)$,

the matrix B is *uniquely identifiable*

Proof idea: overcomplete ICA on "finite noise" transform of $(X_t)_t$

Spotlight: non-Gaussian identifiability – algo./eval.

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Our estimation algorithm:

Mixture of Gaussians as N_t,
variational EM

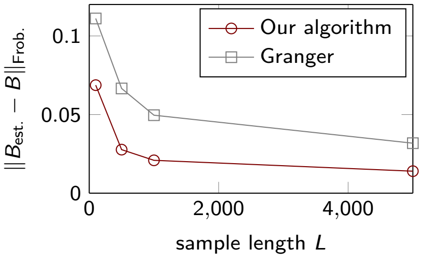
Spotlight: non-Gaussian identifiability – algo./eval.

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Our estimation algorithm:

Mixture of Gaussians as N_t, variational EM

Evaluation on simulated data:



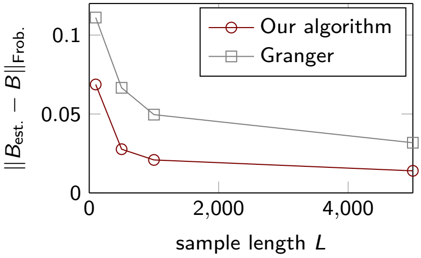
Spotlight: non-Gaussian identifiability – algo./eval.

$$\begin{pmatrix} X_t \\ Z_t \end{pmatrix} := \begin{pmatrix} B & C \\ D & E \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Z_{t-1} \end{pmatrix} + N_t.$$

Our estimation algorithm:

Mixture of Gaussians as N_t, variational EM

Evaluation on simulated data:



Eval. on real economic data:

$$\begin{pmatrix} X^1 \\ X^2 \\ Z \end{pmatrix} = \begin{pmatrix} \text{cheese price} \\ \text{butter price} \\ \text{milk price} \end{pmatrix}$$

$$B^{\text{our}}_{\text{est.}} = \begin{pmatrix} 0.9166 & 0.0513 \\ -0.0094 & 0.9828 \end{pmatrix}$$

$$B^{\text{Granger}}_{\text{est.}} = \begin{pmatrix} 0.8707 & 0.0837 \\ -0.0227 & 0.9559 \end{pmatrix}$$

- Presumed ground truth: no direct eff. between X^k
- Granger more self-consistent (rel. to. complete)

Causal models for decision making via integrative inference – outline

Background in causal models

Motivation and overview over thesis

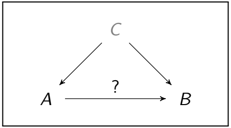
Project 1: Causal inference from confounded time series

Project 2: Causal inference in i.i.d. settings by bounding confounding

Conclusions

Overview: problem and contributions

Goal: infer strength of causal effect of A on B



Problem of hidden confounding again

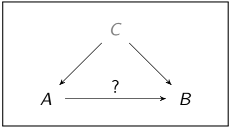
Given:

- $P(A, B)$ (via presumably i.i.d. observations)
- Various forms of additional knowledge

Background: "quasi-experiments" [Shadish et al. 2002] barely using causal models (as introduced above)

Overview: problem and contributions

Goal: infer strength of causal effect of A on B



Problem of hidden confounding again

Given:

- $P(A, B)$ (via presumably i.i.d. observations)
- Various forms of additional knowledge

Background: "quasi-experiments" [Shadish et al. 2002] barely using causal models (as introduced above)

Contributions:

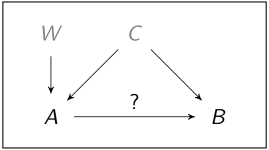
- Six theorems:
  bounds on confounding $\Rightarrow$ bounds on causal effect
- Several prototypical scenarios:
  integration of additional knowledge $\Rightarrow$ bounds on confounding

Spotlight: addressing the ad decision example

Recall: Send out ad letters ($A \in \{0, 1\}$) to influence buying behaviour ($B \in \{0, 1\}$)?

Spotlight: addressing the ad decision example

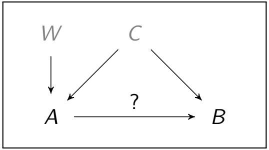
Recall: Send out ad letters ($A \in \{0, 1\}$) to influence buying behaviour ($B \in \{0, 1\}$)?



Here, assume that in a past situation:

- only $P(A, B)$ was recorded – and, say, $\mathrm{I}(A : B) = 0.75$
- C: unmeasured recommendation based on prelim. guidelines
- W: decision A was based on C ($W{=}1$) or done randomly
- have a rough estimate of $P(W)$ – say $P(W) = 0.5$

Spotlight: addressing the ad decision example



Recall: Send out ad letters ($A \in \{0,1\}$) to influence buying behaviour ($B \in \{0,1\}$)?

Here, assume that in a past situation:

- only $P(A,B)$ was recorded – and, say, $\mathrm{I}(A:B) = 0.75$
- C: unmeasured recommendation based on prelim. guidelines
- W: decision A was based on C ($W{=}1$) or done randomly
- have a rough estimate of $P(W)$ – say $P(W) = 0.5$

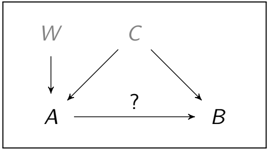
Our results allow to integrate this information:

E.g.: for $\mathfrak{C}_{A\to B}$, a measure of causal influence from A on B,

$\mathfrak{C}_{A\to B}$

Spotlight: addressing the ad decision example

Recall: Send out ad letters ($A \in \{0,1\}$) to influence buying behaviour ($B \in \{0,1\}$)?



Here, assume that in a past situation:

- only $P(A,B)$ was recorded – and, say, $\mathrm{I}(A:B) = 0.75$
- C: unmeasured recommendation based on prelim. guidelines
- W: decision A was based on C ($W{=}1$) or done randomly
- have a rough estimate of $P(W)$ – say $P(W) = 0.5$

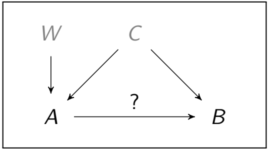
Our results allow to integrate this information:

E.g.: for $\mathfrak{C}_{A \to B}$, a measure of causal influence from A on B,

$\mathfrak{C}_{A \to B} \geq \mathrm{I}(A{:}B) - \mathrm{I}(C{:}A)$

Spotlight: addressing the ad decision example

Recall: Send out ad letters ($A \in \{0,1\}$) to influence buying behaviour ($B \in \{0,1\}$)?



Here, assume that in a past situation:

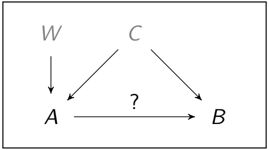
- only $P(A, B)$ was recorded – and, say, $\mathrm{I}(A : B) = 0.75$
- C: unmeasured recommendation based on prelim. guidelines
- W: decision A was based on C ($W{=}1$) or done randomly
- have a rough estimate of $P(W)$ – say $P(W) = 0.5$

Our results allow to integrate this information:

E.g.: for $\mathfrak{C}_{A\to B}$, a measure of causal influence from A on B,

$$\mathfrak{C}_{A\to B} \geq \mathrm{I}(A{:}B) - \mathrm{I}(C{:}A) \overset{\text{corrected}}{\geq} \mathrm{I}(A{:}B) - \log_2(|A|)P(W{=}1)$$

Spotlight: addressing the ad decision example



Recall: Send out ad letters ($A \in \{0, 1\}$) to influence buying behaviour ($B \in \{0, 1\}$)?

Here, assume that in a past situation:

- only $P(A, B)$ was recorded – and, say, $\mathrm{I}(A : B) = 0.75$
- C: unmeasured recommendation based on prelim. guidelines
- W: decision A was based on C ($W{=}1$) or done randomly
- have a rough estimate of $P(W)$ – say $P(W) = 0.5$

Our results allow to integrate this information:

E.g.: for $\mathfrak{C}_{A \to B}$, a measure of causal influence from A on B,

$$\mathfrak{C}_{A \to B} \geq \mathrm{I}(A{:}B) - \mathrm{I}(C{:}A) \overset{\text{corrected}}{\geq} \mathrm{I}(A{:}B) - \log_2(|A|)P(W{=}1) \geq 0.25$$

Conclusions

Summary: causal inference with focus on *integration* of additional knowledge, *approximations*, and *decision* applications

Conclusions

Summary: causal inference with focus on *integration* of additional knowledge, *approximations*, and *decision* applications

Accomplishments – and potential future work:

- Project 1: time helps for removing hidden confounders (barely studied before) – future: beyond linear

Conclusions

Summary: causal inference with focus on *integration* of additional knowledge, *approximations*, and *decision* applications

Accomplishments – and potential future work:

- Project 1: time helps for removing hidden confounders (barely studied before) – future: beyond linear
- Project 2: integrating knowledge that bounds confounding – future: expand results together with domain experts

Conclusions

Summary: causal inference with focus on *integration* of additional knowledge, *approximations*, and *decision* applications

Accomplishments – and potential future work:

- Project 1: time helps for removing hidden confounders (barely studied before) – future: beyond linear
- Project 2: integrating knowledge that bounds confounding – future: expand results together with domain experts
- Project 3 ($\notin$ talk): results driven by cloud computing decision problems – future: more sophisticated experiments

References

C. W. J. Granger. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". English. In: *Econometrica* 37.3 (1969), pp. 424–438. ISSN: 00129682.

J. Pearl. *Causality*. Cambridge University Press, 2000.

William R Shadish, Thomas D Cook, and Donald T Campbell. *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company, 2002.