Seminar in Directed Graphical Models and Causality

2. Equivalence of Factorization and Markov Properties and Independence Testing

Grigor Keropyan

Technical University of Munich

October 27, 2021

Outline

Causal Modeling

1. We observe $\mathbf{X} = (X_1, \dots, X_m)$ random vector
2. $\mathcal{G} = (V, \mathcal{E})$ associated Directed Acyclic Graph (DAG), with $V = 1, \dots, m$
3. Factorization according to DAG $\mathcal{G}$ is
   $f(\mathbf{x}) = \prod_{v \in V} f_{X_v | X_{pa(v)}}(x_v | x_{pa(v)})$

D-separation

In a DAG $\mathcal{G} = (\mathbf{V}, \mathcal{E})$, a path between i and j is **blocked** by $\mathbf{S} \subset \mathbf{V} \backslash \{i, j\}$ whenever there is a node k in the path and one of the following holds:

1. $k \in \mathbf{S}$ and k is not a collider in the path, or
2. $k \notin \mathbf{S}$ and k is a collider in the path and $\forall l \in \mathbf{DE}_k^{\mathcal{G}} \implies l \notin \mathbf{S}$.

Given disjoint subsets $\mathbf{A}, \mathbf{B}, \mathbf{C}$, we say $\mathbf{A}$ and $\mathbf{B}$ are **d-separated** by $\mathbf{C}$ if every path between nodes in $\mathbf{A}$ and $\mathbf{B}$ is blocked by $\mathbf{C}$.

Markov Property

The joint distribution $\mathcal{L}(\mathbf{X})$ of $\mathbf{X}$ is said to be **Markov with respect to the DAG** $\mathcal{G}$ if

$$\mathbf{A}, \mathbf{B} \text{ d-sep. by } \mathbf{C} \implies \mathbf{A} \perp\!\!\!\perp \mathbf{B} | \mathbf{C}.$$

for all disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V}$.

Faithfulness and Causal Minimality

1. We say $\mathcal{L}(\mathbf{X})$ is **faithful to the DAG** $\mathcal{G}$ if

$$\mathbf{A}, \mathbf{B} \text{ d-sep. by } \mathbf{C} \impliedby \mathbf{A} \perp\!\!\!\perp \mathbf{B} | \mathbf{C}.$$

   for all disjoint sets $\mathbf{A}, \mathbf{B}, \mathbf{C} \subseteq \mathbf{V}$.

2. A distribution satisfies **causal minimality** with respect to graph $\mathcal{G}$ if it is Markov with respect to $\mathcal{G}$, but not to any proper subgraph of $\mathcal{G}$.

Markov Equivalence

1. $\mathcal{M}(\mathcal{G}) := \{\mathcal{L}(\mathbf{X}) : \mathcal{L}(\mathbf{X}) \text{ is Markov w.r.t. } \mathcal{G}\}$
2. Two DAGs $\mathcal{G}_1$ and $\mathcal{G}_2$ are called **Markov equivalent** if $\mathcal{M}(\mathcal{G}_1) = \mathcal{M}(\mathcal{G}_2)$.
3. This holds if and only if $\mathcal{G}_1$ and $\mathcal{G}_2$ satisfy same set of d-separations.

Markov Equivalence

Three nodes i, j, k are called **immorality** or **v-structure** if one of them, say j is a child of the others and these parents are not adjacent: $i \to j$, $k \to j$ and $(k, i) \notin \mathcal{E}, (i, k) \notin \mathcal{E}$

Lemma 2 ((Verma and Pearl, 1990))
Two DAGs are Markov equivalent if and only if they have the same v-structures and same skeleton.

Structural Equation Model (SEM)

A structural equation model (SEM) is defined as a tuple $(\mathcal{S}, \mathcal{L}(\mathbf{N}))$, where $\mathcal{S} = (S_1, \ldots, S_p)$ is a collection of p equations

$$S_j : \quad X_j = f_j(\mathbf{PA}_j, N_j), \quad j = 1, \ldots, p \tag{1}$$

Identifiability of SEM

Proposition 1 (Proposition 9 of (Peters et al., 2014))
Let the joint distribution $\mathcal{L}(\mathbf{X})$ of $\mathbf{X} = (X_1, \ldots, X_p)$ is Markov with respect to $\mathcal{G}$ and it has positive density with respect to Lebesgue measure. Then there exists an SEM with graph $\mathcal{G}$ that generates the distribution $\mathcal{L}(\mathbf{X})$.

Post-Nonlinear causal model

Post-Nonlinear (PNL) causal model is a SEM $(\mathcal{S}, \mathcal{L}(\mathbf{N}))$ with equations

$$S_j : \quad X_j = f_{j,2}(f_{j,1}(\mathbf{PA}_j) + N_j), \quad j = 1, \dots, p \tag{2}$$

where $\mathbf{PA}_j$ are the parents of X_j. $N_j \perp\!\!\!\perp \mathbf{PA}_j$ and are jointly independent.

Note that in the case of $f_{j,2}$ $\forall j \in [1, p]$ functions are identity, the PNL causal model becomes Additive Noise Models (ANM).

Identifiabilty of Bivariate PNL Models

Suppose

$$X_2 = f_2(f_1(X_1) + N_2) \text{ and } X_1 = g_2(g_1(X_2) + N_1), \tag{3}$$

where f_2, g_2 are invertible and f_1, g_1 are non-constant functions. Further assume that $p_{N_2}(n_2)$ is positive on $(-\infty, +\infty)$. Then, for every (x_1, x_2) such that $\eta_2^{''} h^{'} \neq 0$ we have

$$\eta_1^{'''} - \frac{\eta_1^{''} h^{''}}{h^{'}} = \left(\frac{\eta_2^{'} \eta_2^{'''}}{\eta_2^{''}} - 2\eta_2^{''}\right) h^{'} h^{''} - \frac{\eta_2^{'''}}{\eta_2^{''}} h^{'} \eta_1^{''} + \eta_2^{'}\left(h^{'''} - \frac{(h^{''})^2}{h^{'}}\right), \tag{4}$$

and

$$\frac{1}{h_1^{'}} = \frac{\eta_1^{''} + \eta_2^{''}(h^{'})^2 - \eta_2^{'} h^{''}}{\eta_2^{''} h^{'}}, \tag{5}$$

where $T_1 := g_2^{-1}(X_1), Z_2 := f_2^{-1}(X_2)$ and
$h := f_1 \circ g_2, h_1 := g_1 \circ f_2, \eta_1(t_1) := \log p_{T_1}(t_1), \eta_2(n_2) := \log p_{N_2}(n_2)$

Non Identifiable Models

N_2	T_1	h
Gaussian	Gaussian	linear
log-mix-lin-exp	log-mix-lin-exp	linear
log-mix-lin-exp	one-sided asymptotically exponential (but not log-mix-lin-exp)	h is strictly monotonic and $h' \to 0$, as $t_1 \to +\infty$ or as $t_1 \to -\infty$
log-mix-lin-exp	generalized mixture of two exponentials	same as above
generalized mixture of two exponentials	two-sided asymptotically exponential	same as above

Table 1: Non identifiable bivariate PNL causal models

Identifiabilty of Multivariate PNL Models

Lemma 2 (Lemma 2 of (Peters et al., 2011) or Lemma 36 of (Peters et al., 2014))
Let Y, Q, N, R be random vectors with continuous joint density $p_{Y,Q,N,R}(y, q, n, r)$. Let $X := f(Y, Q, N)$, where $f : \mathcal{Y} \times \mathcal{Q} \times \mathcal{N} \to \mathcal{X}$ be a measurable function. If $N \perp\!\!\!\perp (Y, Q, R)$ then for all $q \in \mathcal{Q}, r \in \mathcal{R}$ with $p_{Q,R}(q, r) > 0$ and $\bar{Y}_{q,r} := (Y|Q = q, R = r)$, $\bar{X}_{q,r} := (X|Q = q, R = r)$ we have:

$$\bar{X}_{q,r} =_d f(\bar{Y}_{q,r}, q, N).$$

Identifiabilty of Multivariate PNL Models

Lemma 3 (Lemma 37 of (Peters et al., 2014))

Let $\mathcal{L}(\mathbf{X})$ be generated according to SEM as in (1) with corresponding DAG $\mathcal{G}$ and let $X \in \mathbf{X}$.

If $\mathbf{S} \subseteq \mathbf{ND}_X^{\mathcal{G}}$ then $N_X \perp\!\!\!\perp \mathbf{S}$.

Identifiabilty of Multivariate PNL Models

Lemma 4 (Lemma 38 of (Peters et al., 2014))
Let the joint distribution $\mathcal{L}(\mathbf{X})$ of random vector $\mathbf{X}$ has a positive density with respect to Lebesgue measure and is Markov with respect to graph $\mathcal{G}$. Then $\mathcal{L}(\mathbf{X})$ satisfies causal minimality with respect to graph $\mathcal{G}$ if and only if $\forall A \rightarrow B$ in $\mathcal{G}$ and $\forall \mathbf{S} \subset \mathbf{X}$ with $\mathbf{PA}_B^{\mathcal{G}} \setminus \{A\} \subseteq \mathbf{S} \subseteq \mathbf{ND}_B^{\mathcal{G}} \setminus \{A\}$ we have

$$B \not\perp\!\!\!\perp A | \mathbf{S}.$$

Identifiabilty of Multivariate PNL Models

Lemma 5

Given a DAG $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ *and any non adjacent nodes* L *and* W *in* $\mathbf{V}$*. Then, for any set of nodes* $\mathbf{R}$ *in* $\mathbf{V}$ *such that* $\mathbf{R} \subset \mathbf{ND}_W^{\mathcal{G}}$*, then*

$$L, W \text{ d-sep. by } \mathbf{S} \cup \mathbf{R},$$

where $\mathbf{S} := \mathbf{PA}_L^{\mathcal{G}} \cup \mathbf{PA}_W^{\mathcal{G}}$.

Identifiabilty of Multivariate PNL Models

Proposition 6 (Proposition 29 of (Peters et al., 2014))

Let $\mathcal{G}$ and $\mathcal{G}'$ be two different DAGs over variables $\mathbf{X}$.

(i) Suppose the joint distribution $\mathcal{L}(\mathbf{X})$ has a positive density and satisfies the Markov property and causal minimality with respect to $\mathcal{G}$ and $\mathcal{G}'$. Then there exist variables $L, Y \in \mathbf{X}$ such that for the sets $\mathbf{Q} := \mathbf{PA}_L^{\mathcal{G}} \setminus \{Y\}$, $\mathbf{R} := \mathbf{PA}_Y^{\mathcal{G}'} \setminus \{L\}$ and $\mathbf{S} := \mathbf{Q} \cup \mathbf{R}$ we have

- $Y \to L$ in $\mathcal{G}$ and $L \to Y$ in $\mathcal{G}'$
- $\mathbf{S} \subseteq \mathbf{ND}_L^{\mathcal{G}} \setminus \{Y\}$ and $\mathbf{S} \subseteq \mathbf{ND}_Y^{\mathcal{G}'} \setminus \{L\}$

(ii) In particular, if $\mathcal{L}(\mathbf{X})$ is Markov and faithful with respect to $\mathcal{G}$ and $\mathcal{G}'$, then there are variables L and Y such that

- $Y \to L$ in $\mathcal{G}$ and $L \to Y$ in $\mathcal{G}'$
- $\mathbf{PA}_L^{\mathcal{G}} \setminus \{Y\} = \mathbf{PA}_Y^{\mathcal{G}'} \setminus \{L\}$

Proof Idea of Proposition 6

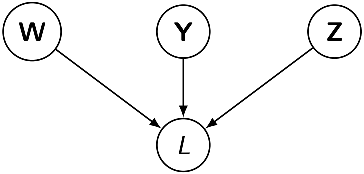

Figure 1: Nodes adjacent to L in $\mathcal{G}$

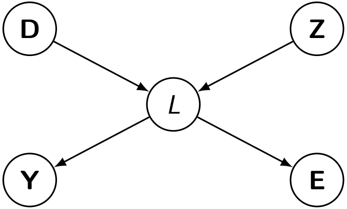

Figure 2: Nodes adjacent to L in $\mathcal{G}'$

Proof Idea of Proposition 6

1. $\mathbf{W} \cup \mathbf{Y} = \emptyset$
    1.1 $\exists D \in \mathbf{D}$
    1.2 $\mathbf{D} = \emptyset$ and $\exists E \in \mathbf{E}$
2. $\mathbf{W} \cup \mathbf{Y} \neq \emptyset$:
   $\exists T \in \mathbf{W} \cup \mathbf{Y}$ such that $\mathbf{DE}_T^{\mathcal{G}'} \cap (\mathbf{W} \cup \mathbf{Y}) = \emptyset$
    2.1 $T \in \mathbf{W}$
    2.2 $T \in \mathbf{Y}$

Identifiabilty of Multivariate PNL Models

Definition 7 (Restricted PNL causal models) (PNL version of Definition 27 of (Peters et al., 2014))

PNL causal model with p variables is called restricted if $\forall j \in \mathbf{V}, i \in \mathbf{PA}_j$ and for all sets $\mathbf{S} \subseteq \mathbf{V}$ with $\mathbf{PA}_j \setminus \{i\} \subseteq \mathbf{S} \subseteq \mathbf{ND}_j \setminus \{i\}$, there is an $x_{\mathbf{S}}$ with $p_{\mathbf{S}}(x_{\mathbf{S}}) > 0$, such that the model

$$X_j^* = f_{j,2}(f_{j,1}(x_{\mathbf{PA}_j \setminus \{i\}}, X_i) + N_j)$$

with a joint distribution $\mathcal{L}(X_i | X_{\mathbf{S}} = x_{\mathbf{S}}, X_j)$ is identifiable, where $X_j^* := X_j | X_{\mathbf{S}} = x_{\mathbf{S}}$.

Identifiabilty of Multivariate PNL Models

Theorem 8 (PNL restricted version of Theorem 28 of (Peters et al., 2014))
Let joint distribution $\mathcal{L}(\mathbf{X})$ of $\mathbf{X}$ is positive and generated by a restricted PNL causal model with graph $\mathcal{G}$ and $\mathcal{L}(\mathbf{X})$ satisfies causal minimality with respect to $\mathcal{G}$. Then, $\mathcal{G}$ is identifiable from the joint distribution.

Multidimensional PNL causal model

Theorem 9

Suppose

$$\mathbf{X}_2 = f_2(f_1(\mathbf{X}_1) + \mathbf{N}_2) \quad \mathbf{X}_1 = g_2(g_1(\mathbf{X}_2) + \mathbf{N}_1), \tag{6}$$

where f_2, g_2 are invertible and f_1, g_1 are non-constant functions. Further assume that $p_{\mathbf{N}}(\mathbf{n})$ is positive on $\mathbb{R}^m$ and the functions $p_{\mathbf{T}}$, $p_{\mathbf{N}}$, f_1, f_2, g_1, g_2 have all second-order partial derivatives. Then for all $i, j \in [1, m]$ and $i \neq j$ the following is true

$$\begin{aligned} &\sum_{l,u=1}^{m} \frac{\partial^2 \eta_1(\mathbf{t})}{\partial \mathbf{t}_j \partial \mathbf{t}_u} \cdot \frac{\partial \mathbf{t}_u}{\partial \mathbf{z}_i} + \sum_{l,u=1}^{m} \frac{\partial^2 \eta_2(\mathbf{n}_2)}{\partial \mathbf{n}_{2,l} \partial \mathbf{n}_{2,u}} \cdot \frac{\partial \mathbf{n}_{2,l}}{\partial \mathbf{n}_{1,j}} \cdot \frac{\partial \mathbf{n}_{2,u}}{\partial \mathbf{z}_i} \\ &\qquad + \sum_{l=1}^{m} \frac{\partial \eta_2(\mathbf{n}_2)}{\partial \mathbf{n}_{2,l}} \cdot \frac{\partial^2 \mathbf{n}_{2,l}}{\partial \mathbf{n}_{1,j} \partial \mathbf{z}_i} = 0. \end{aligned} \tag{7}$$

Thank you!

references

Peters, J., Mooij, J., Janzing, D., and Schölkopf, B. (2011). Identifiability of causal graphs using functional models. pages 589–598, Corvallis, OR, USA. AUAI Press.

Peters, J., Mooij, J. M., Janzing, D., and Schölkopf, B. (2014). Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(58):2009–2053.

Verma, T. and Pearl, J. (1990). Equivalence and synthesis of causal models. In *UAI*.