Undirected Graphs for Causal Structure Learning

Alex Markham
(pronouns: they/them)

Postdoc in Math of Data and AI
Department of Mathematics
KTH Royal Institute of Technology

February 17, 2022
Causality Seminar

Main Idea

- Undirected graphs (usually) offer less powerful/detailed representations of causal structure than DAGs or mixed graphs...
- But they are much easier to work with...
- ...so use undirected graphs first to make the problem easier, then go back to DAGs or mixed graphs

Outline

Undirected Graphs for Latent-induced Dependence

- Undirected graphs (usually) offer less powerful/detailed representations of causal structure than DAGs or mixed graphs...
- Question: when are undirected representations not less powerful/detailed?
- One answer: when considering latent-induced dependence.

Undirected Graphs for Latent-induced Dependence

- The familiar Causal Markov Assumption implies causal sufficiency,i.e., every probabilistic dependence between two observed variables is due to either (i) one of them (indirectly) causing the other, or (ii) both of them having a common cause among the *observed* variables.
- This assumption can be relaxed, allowing the common cause to be unobserved or *latent*.
- We consider a strengthening of this relaxed assumption, namely, what if *all* dependence between the observed variables is because of common latent causes?

Undirected Graphs for Latent-induced Dependence

In this setting, called *strong causal insufficiency*, it turns out that:

- an undirected graph can completely describe the causal structure/conditional independence relations among the observed variables
- furthermore, the undirected graph (via a minimal edge clique cover) facilitates learning an explicit DAG representation of the minimal latent causal model capable of generating the observed data.

Intuition: allows for a causally-interpretable factor analysis

A. Markham and M. Grosse-Wentrup (2020). "Measurement Dependence Inducing Latent Causal Models". In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. PMLR

Undirected Graphs for Causal Kernel Embeddings

- Question: How to tell if samples generated by different underlying causal structures?
- One solution: Project samples into a causally-interpretable space and measure distance/similarity between the points there
- ...then how do we construct a feature map to such a space?
- ...and how do we measure distance there?

Definition (Dependence Contribution Map, $\varphi(\cdot)$)

Let $S \in \mathbb{R}^{n,m}$ be a set of n samples from the vector of m random variables. Let $D \in \mathbb{R}^{n,n,m}$ denote the 3-dimensional array of stacked pairwise distance matrices. Let $C \in \mathbb{R}^{n,n,m}$ to denote these same distance matrices after being doubly-centered. Now standardize the doubly-centered distances to get $Z_{i,i',j} := \frac{C_{i,i',j}}{\bar{D}_{\cdot,\cdot,j}}$. Finally, the dependence contribution map, $\varphi : \mathbb{R}^m \to \mathbb{R}^{m,m}$, is defined as

$$\varphi(S_{i,\cdot}) := Z_{i,\cdot,\cdot}^{\top} Z_{i,\cdot,\cdot} - \mathcal{T}(\alpha).$$

- Same result as pairwise nonlinear (unconditional) independence tests using distance covariance with significance level α.

Definition (Dependence Contribution Kernel, $\kappa(\cdot,\cdot)$)

Let $S, Z, \mathcal{T}$, and φ be as in Definition 1. We define the dependence contribution kernel using the Frobenius (denoted by the subscript $_\mathrm{F}$) inner product and norm: $\kappa(S_{i,\cdot}, S_{i',\cdot}) = \frac{\langle \varphi(S_{i,\cdot}), \varphi(S_{i',\cdot}) \rangle_\mathrm{F}}{\left\|\varphi(S_{i,\cdot})\right\|_\mathrm{F} \left\|\varphi(S_{i',\cdot})\right\|_\mathrm{F}}$.

- Using kernel trick, we can directly calculate similarity (and its inverse, distance) between samples without having to explicitly (and computationally expensively) project them using the feature map.

Undirected Graphs for Causal Kernel Embeddings

Theorem

Let $S \in \mathbb{R}^{n,m}$, $S' \in \mathbb{R}^{n',m}$ be sets of n, n' iid samples drawn respectively from the random variables $X = (X_1, \ldots, X_m)$ and $X' = (X'_1, \ldots, X'_m)$ with finite first moments. Then,

$$\sum_{i=1}^{n} \sum_{i'=1}^{n'} \kappa(S_{i,\cdot}, S'_{i',\cdot}) < 0$$

$$\implies \exists j, j' \in \{1, \ldots, m\} \text{ such that } \mathcal{I}(X_j, X_{j'}; \emptyset) \neq \mathcal{I}(X'_j, X'_{j'}; \emptyset)$$

i.e., X and X' have different unconditional independence relations and therefore different causal structures.

Undirected Graphs for Causal Kernel Embeddings

- In other words, the dependence contribution kernel κ gives us a statistically consistent estimator of when two sets of samples come from different causal structures (up to unconditional equivalence).

- Even stronger (and more specifically), we establish an isometry between distance in the kernel space (quotiented into orthants) and distance in the space of causal ancestral graphs (quotiented into undirected graphs)

A. Markham, R. Das, and M. Grosse-Wentrup (2022). "A Distance Covariance-based Kernel for Nonlinear Causal Clustering in Heterogeneous Populations". In: *Proceedings of the 1st Conference on Causal Learning and Reasoning (CLeaR)*. PMLR

Undirected Graphs for Greedy Equivalence Search

"Optimal Structure Identification with Greedy Search" (Chickering, 2002) introduced the GES algorithm:

- start with empty graph
- greedily add edges, thus moving between Markov equivalence classes (MECs) represented by essential graphs
- when no edge additions improve score, greedily remove edges till arriving at optimal structure

One key insight is to partition DAG space into smaller MEC space, and perform search there; another is the transformational characterization of the MEC.

Definition

The unconditional dependence graph (UDG) of a DAG $\mathcal{G} = (V, E)$ is the undirected graph $\mathcal{U}^{\mathcal{G}} = (V, \{\{v, w\} : v \not\perp_{\mathcal{G}} w\})$.

- Like essential graphs, these form an equivalence class over DAGs (indeed, as shown in the kernel paper, over ancestral graphs), called the unconditional equivalence class (UEC).
- The UEC also admits a transformational characterization.
- It is a partition coarsening of MEC space.
- The UEC can be estimated efficiently from data using, e.g., covariance matrix.

Undirected Graphs for Greedy Equivalence Search

This all leads to the Greedy Unconditional Equivalence Search (GUES) algorithm:

- start with unconditional independence tests to estimate $\mathcal{U}$
- initialize a maximal essential graph in the UEC defined by $\mathcal{U}$
- greedily move (by removing or reversing specific edges) between the essential graphs (MECs) within the UEC till arriving at optimal structure

A. Markham, D. Deligeorgaki, P. Misra, and L. Solus (2022). "Causal Structure Learning with Greedy Unconditional Equivalence Search". In: *Preprint*. on arXiv soon!

Thanks!

Questions?

alex.markham@causal.dev