Joint Probabilistic Inference for Causal Structure Discovery

Dhanya Sridhar and **Lise Getoor**
Computer Science Dept.
University of California Santa Cruz
{dsridhar, getoor}@soe.ucsc.edu

Causal directed acyclic graphical models (DAGs) are powerful reasoning tools in the study and estimation of cause and effect in scientific and socio-behavioral phenomena. In domains where the cause and effect structure is unknown, a key challenge in studying causality with DAGs is learning the structure of causal graphs directly from observational data. Traditional approaches to *causal structure discovery* are categorized as constraint-based or score-based approaches. Score-based methods typically perform greedy search over the space of models whereas constraint-based methods iteratively prune and orient edges using structural and statistical constraints. However, both types of approaches rely on heuristics that introduce false positives and negatives.

Recently, causal structure discovery has been cast as a *MAX-SAT* [1] problem, with d-separation criteria enforcing constraints over assignments to causal edges between variables, similar to constraint-based methods. Jaakkola et al. [2] extend the constraint-satisfaction viewpoint to constrained optimization and formulate a *linear program* to solve the structure discovery problem. In a similar vein, Schmidt et al. [3] use L1-regularized optimization for local Markov blanket identification.

Motivated by these MAX-SAT and optimization-based approaches, we recognize that causal structure discovery can also be viewed as a probabilistic inference problem. We define distributions over model structures and infer the most likely structure given observations, replacing search over structures with optimization. Among several advantages, *joint* probabilistic models need not rely on early pruning or variable orderings, unlike standard constraint-based methods.

Formally, we have a set of n variables $\mathbf{X} = \{X_1 \dots X_n\}$ including latent variables and confounders, and m observations of them given by matrix $\mathbf{O} = \{\mathbf{X}_1 \dots \mathbf{X}_m\}$. We assume that the data come from a true unknown distribution defined by directed acyclic graph (DAG) $\mathcal{G} = (\mathbf{X}, \mathbf{E})$, where the joint probability over $\mathbf{X}$ is a product of local conditional probabilities of each variable given its parents.

In the causal structure inference problem, all possible edges between X_i and X_j correspond to random variables $C_{ij} \in \{0, 1\}$ that are 1 if the directed edge $e_{ij} \in \mathbf{E}$, indicating that X_i causes X_j. Since purely observational data does not suffice to identify a unique true DAG, the output includes undirected edges that indicate association, not causation. Thus, we have random variables $A_{ij} \in \{0, 1\}$ that are 1 when X_i and X_j are dependent. By performing conditional independence tests $I(X_i, X_j | \{X_l \dots X_k\})$ with observations from $\mathbf{O}$, we have *observed* variables I_{ij} that correspond to independence between X_i and X_j.

In the joint probabilistic causal structure inference problem, we want to find

$$\arg\max_{\mathbf{C},\mathbf{A}} P(C_{11}, \dots, C_{nn}, A_{11}, \dots, A_{nn} | I_{ij}) \qquad (1)$$

the *maximum a posteriori* (MAP) assignment to all random variables $\mathbf{C}, \mathbf{A}$ based on a well-defined and tractable joint probability distribution P. However defining such a joint probability distribution is challenging. We would like it to encode the following information:

- D-separation and acyclicity constraints as in traditional constraint-based approaches such as PC or FCI [5, 4]
- Flexible priors and domain knowledge to guide optimization over likely DAG structures
- Information from multiple sources of evidence
- Background information about likely causes, targets and latent variables

The open problem that we pose is how to incorporate all of this information in a unified and scalable probabilistic framework.

References

[1] Antti Hyttinen, Patrik O Hoyer, Frederick Eberhardt, and Matti Jarvisalo. Discovering cyclic causal models with latent variables: A general sat-based procedure. *arXiv preprint arXiv:1309.6836*, 2013.

[2] Tommi Jaakkola, David Sontag, Amir Globerson, and Marina Meila. Learning bayesian network structure using LP relaxations. In *International Conference on Artificial Intelligence and Statistics*, pages 358–365, 2010.

[3] Mark Schmidt, Alexandru Niculescu-Mizil, Kevin Murphy, et al. Learning graphical model structure using l1-regularization paths. In *AAAI*, volume 7, pages 1278–1283, 2007.

[4] Peter Spirtes. An anytime algorithm for causal inference. In *AISTATS*. Citeseer, 2001.

[5] Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1):62–72, 1991.