1. Motivation

Causal models extend purely probabilistic models, enabling reasoning about joint distributions of random variables in the presence of well-defined interventions[2].

Causal probabilistic programs include common programming constructs such as recursion, looping, and conditional branching. Conditional branching[1] can be used to represent context-dependent causal structure, i.e. for some subset of random variables C, there exist two execution paths i and j through the program, such that $P(C_i) \neq P(C_j)$. A set of techniques has been developed over the past 25 years to learn the structure of causal graphical models from observational (non-experimental) data[3], however they implicitly assume that the graph structure is context-independent.

Program P1: Context-Independent Structure

\begin{align*}
A &\leftarrow f_A() \\
B &\leftarrow f_B(A) \\
C &\leftarrow f_C(A) \\
D &\leftarrow f_D(B, C)
\end{align*}

Program P2: Context-Dependent Structure

\begin{align*}
&\text{if } \text{Bernoulli}(p) \text{ then} \\
&\quad A \leftarrow f_A() \\
&\quad C \leftarrow f_C(A) \\
&\quad \text{else} \\
&\quad C \leftarrow f_C() \\
&\quad A \leftarrow f_A(C) \\
&\quad B \leftarrow f_B(A) \\
&\quad D \leftarrow f_D(B, C)
\end{align*}

2. Equivalence

Two models $M$ and $M'$ are observationally equivalent if $P(X|M) = P(X|M')$. Two models $M$ and $M'$ are interventionally equivalent over a set of intervenable random variables $Y$ if $P(X|M, \text{do}(Y = y')) = P(X|M', \text{do}(Y = y')), \forall y' \in \text{domain}(Y)$

$M_1$ and $M_2$ are observationally equivalent given particular conditional probability distributions, but are not interventionally equivalent except for the trivial case of $A \perp B$.

Given a model $M = (G, F)$, where $G = (V, E)$ is a directed graph and $F$ is a set of conditional probability distributions $P(X|Pa(X))\forall X \in V$, there exists a causal probabilistic program which is observationally and interventional equivalent to $M$ over all $X \in V$.

3. Structure Discovery Experiments

We evaluate the performance of graph-based structure discovery algorithms when the generative process is a causal probabilistic program with context-dependent causal structure using synthetic experiments.

To do this we: (1) generate observational samples from programs P1 and P2, (2) learn a Markov equivalence class of graphical models using the max-min hill climbing algorithm[4], (3) non-parametrically estimate local conditional probability distributions, and (4) generate interventional samples from both the causal probabilistic program and the learned graphical model for the intervention $\text{do}(A = a)$.

4. Conclusions

- We demonstrate that simple causal probabilistic programs with conditional branching can represent causal processes that are not learned effectively by the most common algorithms for learning causal graphical models.
- This is despite the fact that the same learning procedures produce observational estimates that closely approximate the programs’ observational distribution.
- The task of learning the structure of causal models with context-dependent causal structure remains an important research frontier.

Acknowledgements — Thanks to Javier Barrientos, Raleigh Clary, Vikash Mansinghka, and the members of the MIT probabilistic computing project for thoughtful contributions and comments. This material is based upon work supported by the United States Air Force under Contract No. FA8750-17-C-0120. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force.