Causal Graphs vs. Causal Programs
The Case of Conditional Branching

Sam Witty and David Jensen

Knowledge Discovery Laboratory • College of Information and Computer Sciences • University of Massachusetts Amherst

1. Motivation

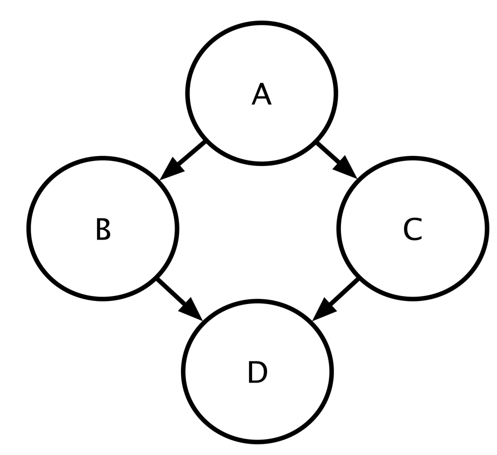
Causal models extend purely probabilistic models, enabling reasoning about joint distributions of random variables in the presence of well-defined interventions[2].

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 $Pa(C_i) \neq Pa(C_j)$.

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

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



Program P2: Context-Dependent Structure

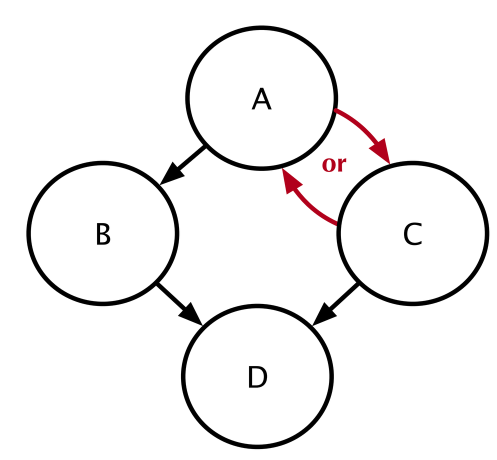
if $Bernoulli(p)$ **then**

$$\begin{aligned} &A \leftarrow f_A() \\ &C \leftarrow f_C(A) \end{aligned}$$

else

$$\begin{aligned} &C \leftarrow f_{C'}() \\ &A \leftarrow f_{A'}(C) \end{aligned}$$

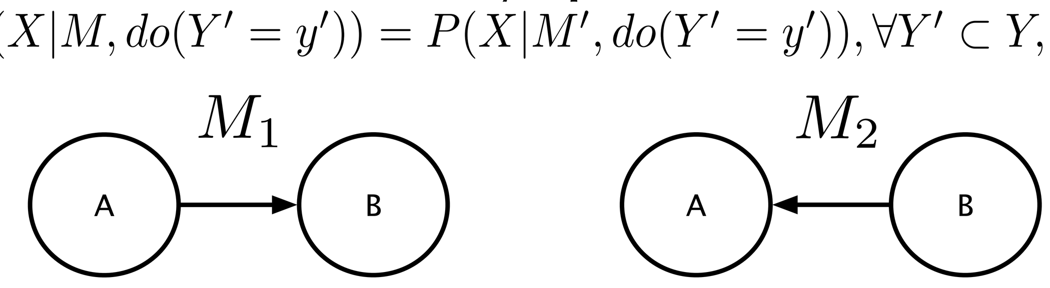
$$\begin{aligned} &B \leftarrow f_B(A) \\ &D \leftarrow f_D(B, C) \end{aligned}$$



2. Equivalence

Two models M and M' are **observationally equivalent** if $P(X|M) = P(X|M')$

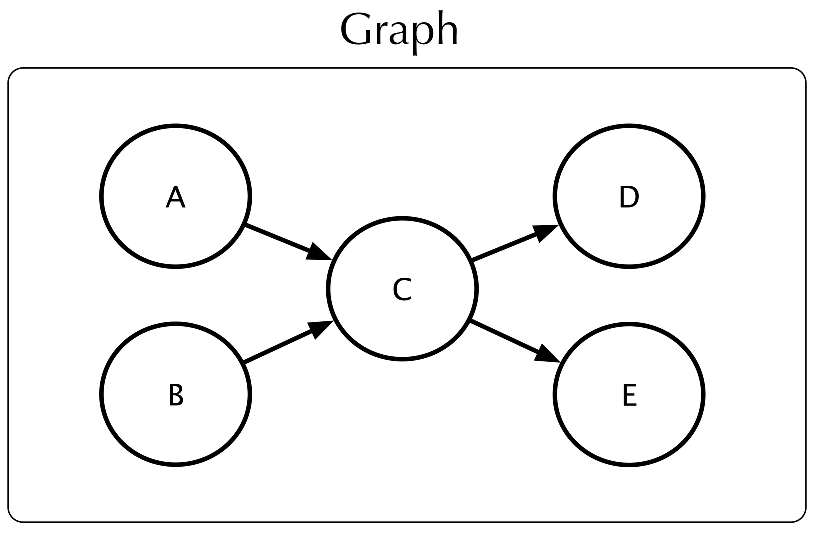
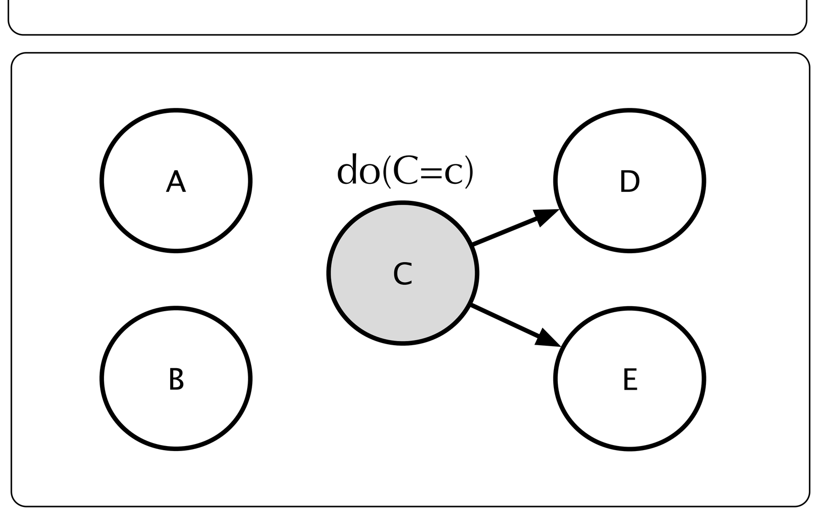
Two models M and M' are **interventionally equivalent** over a set of intervenable random variables Y if $P(X|M, do(Y' = y')) = P(X|M', do(Y' = y')), \forall Y' \subset Y, y' \in domain(Y')$



Interventional

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

> 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 interventionally equivalent to M over all $X \in V$.

	Graph	Program
Observational		$A \leftarrow f_A()$ $B \leftarrow f_B()$ $C \leftarrow f_C(A, B)$ $D \leftarrow f_D(C)$ $E \leftarrow f_E(D)$
Interventional		$A \leftarrow f_A()$ $B \leftarrow f_B()$ $C \leftarrow c$ $D \leftarrow f_D(C)$ $E \leftarrow f_E(D)$

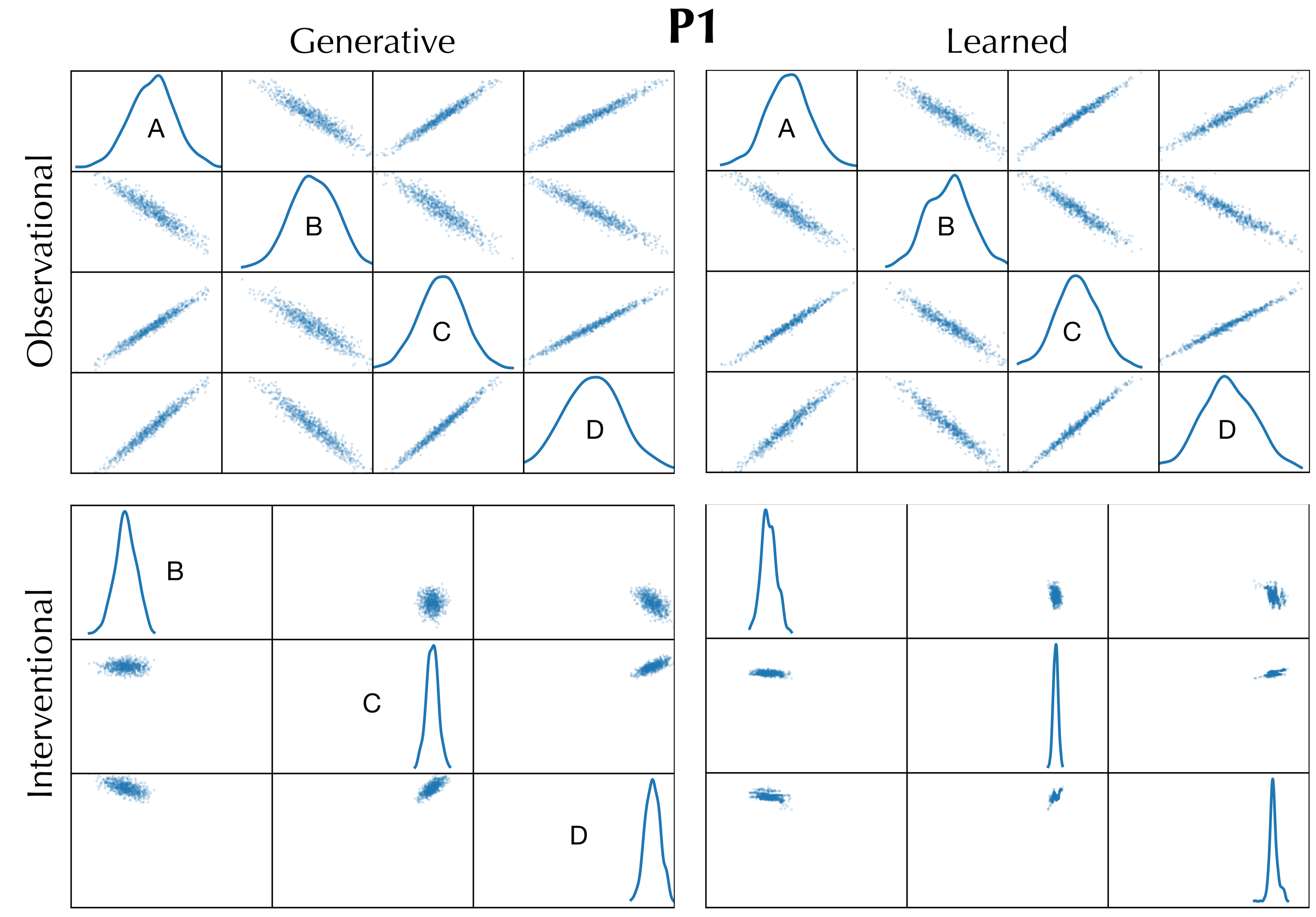
> Given a probabilistic program with context-dependent causal structure there does not exist an interventionally equivalent causal graphical model. However, there may exist an observationally equivalent causal graphical model.

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.

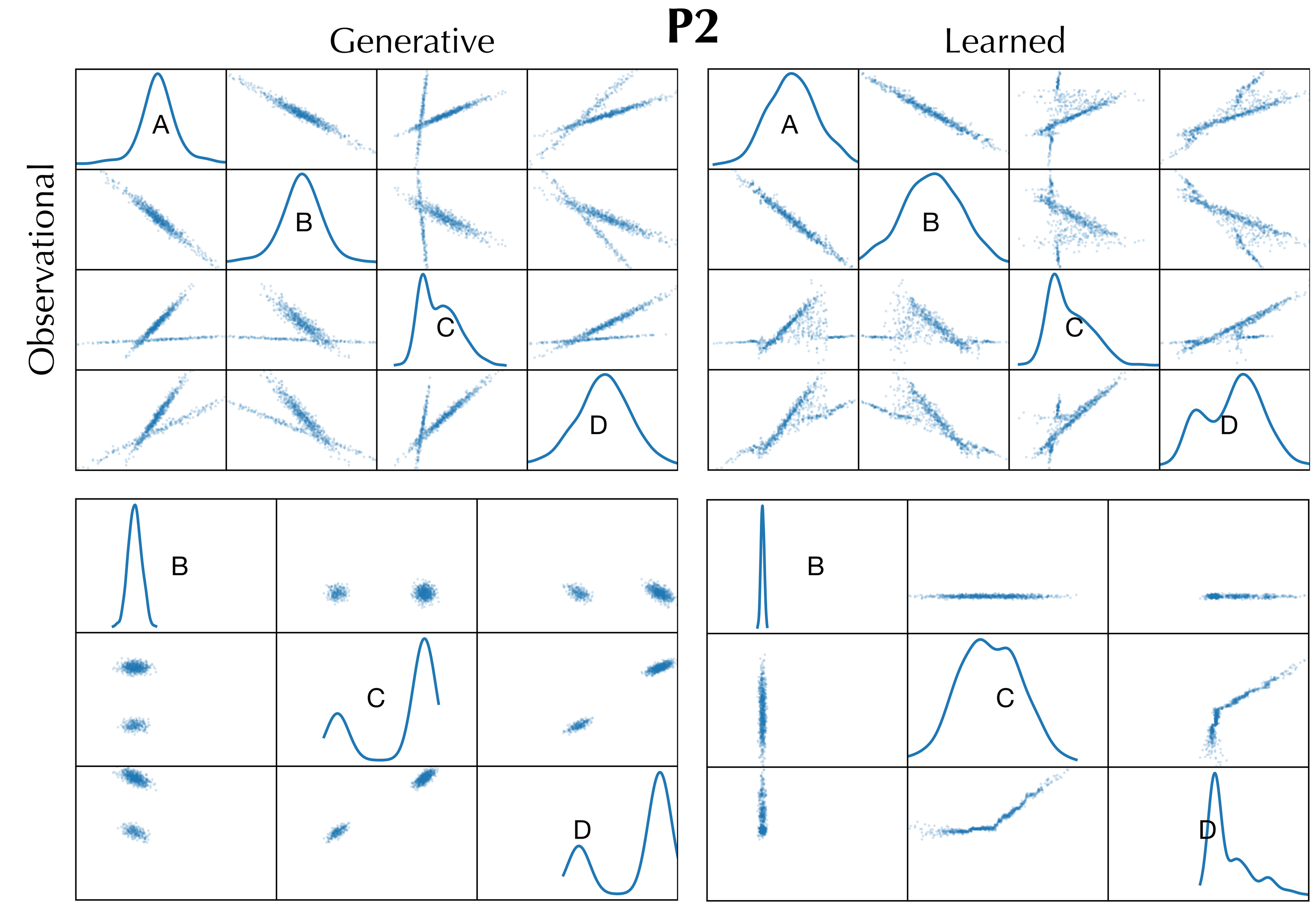
To do this we: (1) generate observational samples from programs P1 and P2, (2) learn a Markov equivalence class of graphical models using the max-min hill climbing algorithm[4], (3) non-parametrically estimate local conditional probability distributions, and (4) generate interventional samples from both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

P1



> When the generative process is probabilistically and interventionally equivalent to a causal graphical model, the graph-based causal discovery procedure produces estimates that closely approximate the observational and interventional distributions.

P2



> When the generative process is not interventionally equivalent to any causal graphical model, the graph-based causal discovery procedure produces estimates that closely approximate the observational distribution, but deviate significantly from the interventional distribution.

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.

Acknowledgments — Thanks to Javier Burroni, Kaleigh 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.

[1] Minka, T. and Winn, J. Gates, NIPS, 2008.
[2] Pearl, J., *Causality.* Cambridge University Press, 2nd edition, 2009.
[3] Spirtes, P., Glymour, C., and Scheines, R. *Causation, Prediction, and Search*. MIT Press, 2nd edition, 2000.
[4] Tsamardinos, I., Brown, L., Aliferis, C. The max-min hill-climbing Bayesian network structure learning algorithm, Machine learning, 2006.