Difference graph over two populations: Implicit Difference Inference algorithm

Daria Bystrova[1] **Emilie Devijver**[1] **Vardan Manucharian**[1] **Julie Mondet**[2,3] **Pascal Mossuz**[2,3]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, F38000, Grenoble, France,
[2]Univ. Grenoble Alpes, Institute for Advanced biosciences, INSERM U1209/CNRS U5039, 38700 Grenoble, France
[3] Institute of Biology and Pathology, Hospital of Grenoble Alpes (CHUGA), CS 20217, 38043 Grenoble, CEDEX 9, France

Abstract

Comparing causal relations across different populations is essential in various fields, including medicine and ecology. Recently, several methods have been developed to directly infer difference graphs from observational data. These methods rely on semi-parametric assumptions and suppose that the data is continuous. We propose a new approach for discovering causal difference graphs without any semi-parametric assumption and can be applied on continuous or discrete or mixed data. We provide theoretical guarantees of the new method and test it on simulated data.

1 INTRODUCTION

Understanding the causal relations among variables is an important topic in many domains, such as medicine [Belyaeva et al., 2021] and ecology [Pellissier et al., 2018]. Causal relationships between variables can be represented by a structural causal model (SCM) [Pearl, 2000]. Each SCM can be graphically represented using a causal graph $\mathcal{G} = (\mathbb{V}, \mathbb{E})$, where the nodes $\mathbb{V}$ represent random variables, and the edges $\mathbb{E}$ represent causal relationships. In this paper, we consider only causal graphs that are directed acyclic graphs (DAGs). In many applications causal graphs are unknown, so causal discovery methods infer causal graphs from observational data. One of the most well-known algorithm for discovering causal graphs is the PC algorithm [Spirtes et al., 2000] which assumes Faithfulness and Causal Sufficiency [Spirtes et al., 2000], but no parametric assumptions on the underlying distribution.

In many applications, researchers are interested in comparing relationships between different groups. However, inferring a causal graph from observational data is already challenging, so discovering two different causal graphs and then comparing them is not optimal. Additionally, such an approach would not account for all types of differences between the groups, because the graphs could be the same but encode different causal strength. Therefore, recently, new methods have emerged to discover directly a difference DAG $\mathcal{D} = (\mathbb{V}, \mathbb{E})$ from observational data, such as DCI [Wang et al., 2018] and iSCAN [Chen et al., 2024]. Both methods assume causal sufficiency, parametric assumptions on the underlying distributions, and that the graphs share the same topological order and they require that the variables are continuous.

In this work, we propose a new algorithm, called Implicit Difference Inference (IDI), which directly infers the difference DAG without any parametric assumptions on the underlying distribution and can be applied to continuous, discrete or mixed data. The proposed method is based on the PC algorithm but does not require the faithfulness assumption, only a lighter version of faithfulness.

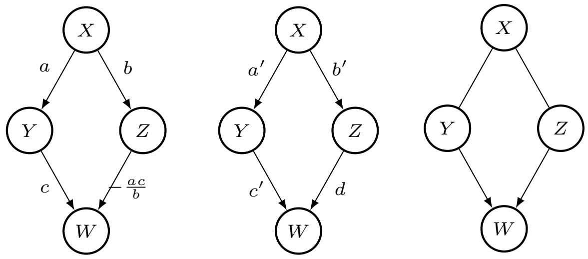
2 SETTING

Consider there exists an unknown true SCM M on the random variables $X_1, \dots, X_p$. Let M^1 and M^2 be two underlying SCMs obtained by a set of interventions (structural or parametric) on M, and let $\mathbb{P}^1$ and $\mathbb{P}^2$ be the underlying distributions of M^1 and M^2, respectively. We have two datasets sampled respectively from the distributions $\mathbb{P}^1$ and $\mathbb{P}^2$. Our goal is to infer the difference graph $\mathcal{D} = (\mathbb{V}, \mathbb{E})$, where the vertices set is $\mathbb{V} = \{X_1, \dots, X_p\}$ and there is an edge between two vertices if there is a significant difference between the two datasets relating those vertices.

The algorithm we present to discover $\mathcal{D}$ is based on the PC algorithm. Instead of using an independence test, we use a statistical test on equality of the dependence measures (we state everything with mutual information, as it is a general form, but any other dependence measure can be used):

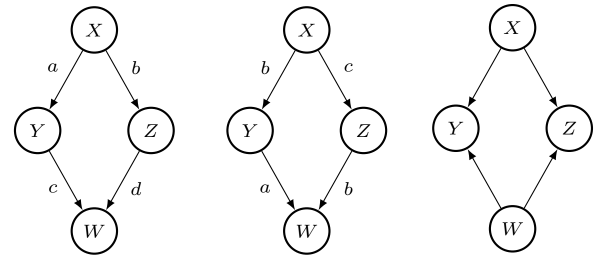
$$H_0 : I_{\mathbb{P}^1}(X_1, X_2|X_3) = I_{\mathbb{P}^2}(X_1, X_2|X_3),$$

for I the mutual information (for estimation and test, see



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$ (c) $\hat{\mathcal{D}}$

Figure 1: Violation of faithfulness but not interventional faithfulness. PC algorithm will not be able to infer $\mathcal{G}_1$, but IDI is able to infer the CPDAG $\hat{\mathcal{D}}$ of the true difference DAG $\mathcal{D}$ (which, in that case, coincides with $\mathcal{D}$).



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$ (c) $\hat{\mathcal{D}}$

Figure 2: Violation of interventional faithfulness. IDI is not able to discover the CPDAG $\hat{\mathcal{D}}$ of the true difference DAG.

[Runge, 2018] for continuous variables and [Zan et al., 2022] for mixed data).

2.1 IMPLICIT DIFFERENCE INFERENCE ALGORITHM

We follow the main steps of the PC algorithm [Spirtes et al., 2000, Colombo and Maathuis, 2014] to construct the difference DAG. First, IDI starts with the complete undirected graph on $\{X_1, \dots, X_p\}$. Then, for every pair of variables, we check if their dependencies are similar within the two datasets or if they can be separated by a separation set. For every size of conditional set (from 0 to $p-2$), we test if the mutual information on the two datasets are equal. As soon as there exists a conditional set such that the mutual information on the two datasets are equal, we remove the edge between the two variables. We use the four rules of the PC algorithm to orient the edges: first, the origin of causality to orient V structures; then, the classical three rules in a loop to orient as many edges as possible. The algorithm is summarized in Algorithm 1 in the Supplementary materials.

2.2 THEORETICAL GUARANTEES

The IDI algorithm is developed to discover difference DAGs rather than causal graphs. Unlike the PC algorithm, IDI does not require the faithfulness assumption, which asserts that if two variables are independent, they remain independent under all possible conditioning sets. This assumption is often criticized for being too stringent, particularly when conditional independencies result from subtle cancellations of causal effects.

To illustrate this, consider the diamond structure introduced in Figure 1. If there exists canceling paths from X to W, the PC algorithm might conclude that X and W are independent, thereby removing the edge between them. However, if the causal coefficients are slightly altered, making each path's strength different, X and W may show different dependencies in the second graph. Therefore, the IDI algorithm would only remove the edge between X and W if it conditions on both Y and Z.

IDI bypasses the faithfulness assumption by not relying on exact cancellations of paths, making it robust in scenarios where the PC algorithm might fail. Similarly, IDI does not require the minimality assumption, which ensures that every edge in the DAG represents a dependence in the distribution. This relaxation allows IDI to handle more complex or over-specified models without needing the causal graph to be in its minimal form. While IDI relaxes some traditional assumptions, it introduces its own set of requirements. Specifically, the IDI algorithm relies on:

Causal Sufficiency: This assumption requires that there are no hidden confounders, meaning all relevant variables are observed.

Interventional Faithfulness: Unique to IDI, this assumption posits that intervening on a vertex will change the mutual information between this node and all its ancestors. This ensures that intervention effects are detectable through changes in the mutual information, enabling the identification of difference DAGs. It implies that the same canceling paths (violations of faithfulness) cannot exist in all causal graphs and prohibits swapping coefficients between causal graphs as in Figure 2. Remark that depending on the dependence measure considered, several implications can be drawn. For example, if we consider the causal graph with two variables X and Y such that $X \to Y$. Suppose that in M^1, $Y = aX + \xi_y$ in M^1, $Y = -aX + \xi_y$, then in this case the two mutual informations are equal. This scenario can be avoided by assuming linearity and replacing mutual information with correlation, but this would imply the need of a parametric assumption.

Assuming the above assumptions given perfect conditional independence information about all pairs of variables, the IDI algorithm discovers the CPDAG $\hat{\mathcal{D}}$ of the true difference DAG $\mathcal{D}$. Experiments are available in the Supplementary materials.

References

Anastasiya Belyaeva, Chandler Squires, and Caroline Uhler. Dci: learning causal differences between gene regulatory networks. *Bioinformatics*, 37(18):3067–3069, 2021.

Tianyu Chen, Kevin Bello, Bryon Aragam, and Pradeep Ravikumar. iscan: Identifying causal mechanism shifts among nonlinear additive noise models. *Advances in Neural Information Processing Systems*, 36, 2024.

Diego Colombo and Marloes H. Maathuis. Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, 15(116):3921–3962, 2014.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-77362-8.

Loïc Pellissier, Camille Albouy, Jordi Bascompte, Nina Farwig, Catherine Graham, Michel Loreau, Maria Alejandra Maglianesi, Carlos J Melián, Camille Pitteloud, Tomas Roslin, et al. Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93 (2):785–800, 2018.

Jakob Runge. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In *International Conference on Artificial Intelligence and Statistics*, pages 938–947. Pmlr, 2018.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.

Yuhao Wang, Chandler Squires, Anastasiya Belyaeva, and Caroline Uhler. Direct estimation of differences in causal graphs. *Advances in neural information processing systems*, 31, 2018.

Lei Zan, Anouar Meynaoui, Charles K Assaad, Emilie Devijver, and Eric Gaussier. A conditional mutual information estimator for mixed data and an associated conditional independence test. *Entropy*, 24(9):1234, 2022.

Difference graph over two populations: Implicit Difference Inference algorithm (Supplementary Material)

Daria Bystrova[1] **Emilie Devijver**[1] **Vardan Manucharian**[1] **Julie Mondet**[2,3] **Pascal Mossuz**[2,3]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, F38000, Grenoble, France,
[2]Univ. Grenoble Alpes, Institute for Advanced biosciences, INSERM U1209/CNRS U5039, 38700 Grenoble, France
[3] Institute of Biology and Pathology, Hospital of Grenoble Alpes (CHUGA), CS 20217, 38043 Grenoble, CEDEX 9, France

A PSEUDOCODE OF THE IDI ALGORITHM

Algorithm 1 IDI algorithm

function IDI($\mathbb{V}$, $\mathbb{P}^1$, $\mathbb{P}^2$):
    Form the complete undirected graph $\hat{\mathcal{D}}$ on the vertex set $\mathbb{V}$.
    $n \leftarrow 0$
    repeat
        repeat
            select an ordered pair of variables X and Y that are adjacent in $\hat{\mathcal{D}}$ such that $Adjacencies(\hat{\mathcal{D}}, X) \setminus Y$ has cardinality greater than or equal to n, and a subset S of $Adjacencies(\hat{\mathcal{D}}, X) \setminus Y$ of cardinality n
            if $I^{\mathbb{P}^1}(X, Y|S) = I^{\mathbb{P}^2}(X, Y|\mathbb{S})$ **then**
                delete edge $X - Y$ from $\hat{\mathcal{D}}$
                record S in $Sepset(X, Y)$ and $Sepset(Y, X)$
            end if
        until all ordered pairs of adjacent variables X and Y such that $Adjacencies(\hat{\mathcal{D}}, X) \setminus Y$ has cardinality greater than or equal to n and all subsets S of $Adjacencies(\hat{\mathcal{D}}, X) \setminus Y$ of cardinality n have been tested
        $n \leftarrow n + 1$
    until for each ordered pair of adjacent vertices X, Y, $Adjacencies(\hat{\mathcal{D}}, X) \setminus Y$ is of cardinality less than n
    For each $X - Z - Y$ in $\hat{\mathcal{D}}$ such that $X - Y$ is not in $\hat{\mathcal{D}}$, if $Z \notin Sepset(X, Y)$ then orient the substructure as $X \rightarrow Z \leftarrow Y$
    repeat
        Apply Rules 1,2,3 of the PC algorithm
    until no edge can be oriented
    return $\hat{\mathcal{D}}$
end function

B EXPERIMENTS

The simulated 100 datasets corresponding to the two causal graphs presented in Figure 1. Each dataset corresponding to $\mathcal{G}^1$ are simulated used the following generator:

$$\begin{aligned}
X &:= \xi^x \\
Y &:= aX + \xi^y \\
Z &:= bX + \xi^z \\
W &:= cY - \frac{ac}{b}Z + \xi^w
\end{aligned} \tag{1}$$

and the datasets corresponding to $\mathcal{G}^2$ are simulated using the following generator:

$$\begin{aligned}
X &:= \xi^x \\
Y &:= a'X + \xi^y \\
Z &:= b'X + \xi^z \\
W &:= c'Y + dZ + \xi^w
\end{aligned} \tag{2}$$

where d is different than $\frac{a'c'}{b'}$.

The results of our experiment showed that IDI was able to recover the CPDAG of true difference DAG in 90% of the datasets.