Algorithmic Causal Inference

Behrad Moniri

Department of Electrical Engineering
Sharif University of Technology

Equivalence of Algorithmic Markov Conditions

Given strings $x_1, \ldots, x_n$ and directed acyclic graph G, the following conditions are equivalent:

1. **Recursive Form (Markov Factorization):**

$$K(x_1, \ldots, x_n) = \sum_{j=1}^{n} K(x_j | pa_j^*)$$

2. **Local Markov Conditions:**

$$I(x_j : nd_j | pa_j^*) = 0$$

3. **Global Markov Conditions:**

$$I(S : T | R) = 0$$

if R d-separates S and T.

Algorithmic Model for Causality

Algorithmic Model for Causality

Let G be a DAG formalizing the causal structure among the strings $x1, \ldots, xn$. Then every x_j is computed by a program q_j with length $O(1)$ from its parents pa_j and an additional independent inputs n_j

$$x_j = q_j(pa_j, n_j).$$

Here, by independence we mean $I(n_j : n^{-j}) = 0$.

Algorithmic Model for Causality

Algorithmic Model for Causality

Let G be a DAG formalizing the causal structure among the strings $x1, \dots, xn$. Then every x_j is computed by a program q_j with length $O(1)$ from its parents pa_j and an additional independent inputs n_j

$$x_j = q_j(pa_j, n_j).$$

Here, by independence we mean $I(n_j : n^{-j}) = 0$.

Note!

the Turing machine simulating the process would not necessarily halt on all inputs pa_j, n_j.

Algorithmic model implies Markov

Algorithmic model implies Markov

Let $x_1, \ldots, x_n$ be generated by the algorithmic model. Then they satisfy the algorithmic Markov condition with respect to G.

Relative Causality

- Consider genetic sequences $s1, s2$ of two persons that are not relatives. There is a significant violation of $I(s1 : s2) = 0$ due to the fact that both genes are taken from humans.

Relative Causality

- Consider genetic sequences $s1, s2$ of two persons that are not relatives. There is a significant violation of $I(s1 : s2) = 0$ due to the fact that both genes are taken from humans.
- For this reason, every discussion on whether there exists a causal link between two objects (or individuals) requires a specification of the background information.

Relative Causality

- Consider genetic sequences $s1, s2$ of two persons that are not relatives. There is a significant violation of $I(s1 : s2) = 0$ due to the fact that both genes are taken from humans.
- For this reason, every discussion on whether there exists a causal link between two objects (or individuals) requires a specification of the background information.
- Causality is relative!

Relative Causality

- Consider genetic sequences $s1, s2$ of two persons that are not relatives. There is a significant violation of $I(s1 : s2) = 0$ due to the fact that both genes are taken from humans.
- For this reason, every discussion on whether there exists a causal link between two objects (or individuals) requires a specification of the background information.
- Causality is relative!

Note!

In the statistical version of the link between causality and dependence, the relevance of the background information is less obvious because it is evident that statistical methods are always applied to a given statistical ensemble.

From Algorithmic to Statistical

- Consider random variables X and Y.
- The machine S, generates samples of X according to $P(X)$.
- M that generates y-values according to $P(Y|X)$.
- If $I\big(P(X) : P(Y|X)\big) >> 0$, there must be a causal link that explains the similarities in the design of S and M.

From Algorithmic to Statistical

- Consider random variables X and Y.
- The machine S, generates samples of X according to $P(X)$.
- M that generates y-values according to $P(Y|X)$.
- If $I\big(P(X) : P(Y|X)\big) >> 0$, there must be a causal link that explains the similarities in the design of S and M.

Postulate

A causal hypothesis G (i.e., a DAG) is only acceptable if the shortest description of the joint density P is given by a concatenation of the shortest description of the Markov kernels, i.e.

$$K(P(X_1, \ldots, X_n)) = \sum_j K(P(X_j|PA_j))$$

From Algorithmic to Statistical

- Assume that you are given two strings **x** and **y** of length n.

From Algorithmic to Statistical

- Assume that you are given two strings $\mathbf{x}$ and $\mathbf{y}$ of length n. You have noted that $\mathbf{x} = \mathbf{y}$.

From Algorithmic to Statistical

- Assume that you are given two strings $\mathbf{x}$ and $\mathbf{y}$ of length n. You have noted that $\mathbf{x} = \mathbf{y}$. Consider to scenarios:

From Algorithmic to Statistical

- Assume that you are given two strings **x** and **y** of length n. You have noted that $\mathbf{x} = \mathbf{y}$. Consider to scenarios:
  1. Each (x_j, y_j) drawn independently drawn from $P(X, Y)$.
  2. **x** and **y** are single instances of string-valued random variables X and Y.

From Algorithmic to Statistical

- Assume that you are given two strings **x** and **y** of length n. You have noted that $\mathbf{x} = \mathbf{y}$. Consider to scenarios:
  1. Each (x_j, y_j) drawn independently drawn from $P(X, Y)$.
  2. **x** and **y** are single instances of string-valued random variables X and Y.
- The difference between (I) and (II) is crucial for statistical causal inference:
  - In case (I), statistical independence is rejected with high confidence.
  - In scenario (II), no evidence for statistical dependence.

$$P(X, Y) = \delta_{\mathbf{x}} \delta_{\mathbf{y}}$$

From Algorithmic to Statistical

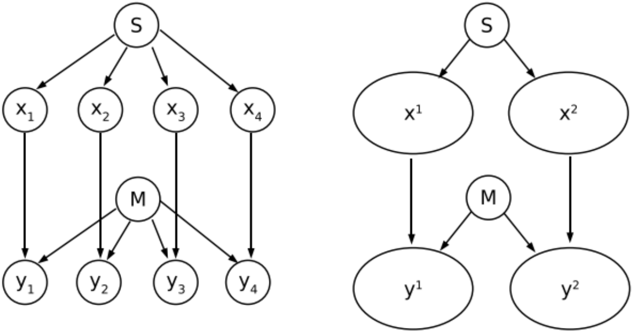
- Assume that you are given two strings **x** and **y** of length n. You have noted that $\mathbf{x} = \mathbf{y}$. Consider to scenarios:
  1. Each (x_j, y_j) drawn independently drawn from $P(X, Y)$.
  2. **x** and **y** are single instances of string-valued random variables X and Y.
- The difference between (I) and (II) is crucial for statistical causal inference:
  - In case (I), statistical independence is rejected with high confidence.
  - In scenario (II), no evidence for statistical dependence.

$$P(X, Y) = \delta_{\mathbf{x}} \delta_{\mathbf{y}}$$

- Algorithmic causal inference, on the other hand, infers a causal link in both cases because the equality $\mathbf{x} = \mathbf{y}$ requires an explanation.

Relation between two scenarios

- Switching between (I) and (II) then consists merely in shifting the causal connection to another level:
  - In the i.i.d setting, every x_j must be causally linked to y_j.
  - In case (II), there must be a connection between the two mechanisms that for instance, be due to the fact that two machines emitting the same string were designed by the same engineer.



$$I(\mathbf{x}^1 : \mathbf{y}^2 | \mathbf{x}^2) = 0 \qquad I(\mathbf{x}^2 : \mathbf{y}^1 | \mathbf{x}^1) = 0$$

The most remarkable property of this, is that they are asymmetric with respect to exchanging the roles of X and Y