A Critique of Inductive Causation

Christian Borgelt and Rudolf Kruse

Department of Knowledge Processing and Language Engineering
Otto-von-Guericke-University of Magdeburg
Universitätsplatz 2, D-39106 Magdeburg, Germany
E-mail: borgelt@iws.cs.uni-magdeburg.de

Abstract: In this paper we consider the problem of inducing causal relations from statistical data. Although it is well known that a correlation does not justify the claim of a causal relation between two measures, the question seems not to be settled. Research in the field of Bayesian networks revived an approach suggested in [16]. It is based on the idea that there are relationships between the causal structure of a domain and its corresponding probability distribution, which could be exploited to infer at least part of the causal structure from a set of dependence and independence statements. This idea was developed into the inductive causation algorithm [14]. We review this algorithm and examine the assumptions underlying it.

1 Introduction

If A causes B, an occurrence of A should be accompanied or (closely) followed by an occurrence of B. That causation implies conjunction is the basis of all reasoning about causation in statistics. But is this enough to infer causal relations from statistical data, and, if not, are there additional assumptions that provide reasonable grounds for such inference? These are the questions we discuss here.

An appropriate framework for such a discussion is the theory of Bayesian networks. Research in this field is influenced from two directions. In the first place, Bayesian networks are studied on purely statistical grounds as one of several approaches to make reasoning in multi-dimensional domains feasible by decomposing the uncertainty information available about the domain [9]. Among such approaches, Bayesian networks [15] and Markov networks [11] are the best known probabilistic methods. Others include the more general valuation-based networks [19] and possibilistic networks [10, 3].

Secondly, Bayesian networks are studied as descriptions of a structure of causal influences. Since they use conditional probability distributions which possess an inherent direction, the idea suggests itself to "direct" the distributions in such a way that they represent the causal influences. Indeed, human experts often start from a causal model of the underlying domain and choose the conditional probability distributions of the Bayesian network accordingly.

Therefore, in Bayesian networks, statistics and causal modeling are conjoined. This is emphasized by the d-separation criterion [15, 4], which allows us to read the probabilistic dependences and independences from the causal structure underlying a Bayesian network. In the sequel an algorithm, the so-called inductive causation algorithm [14], was suggested to invert this procedure and to infer at

least part of the causal structure from observed dependences and independences. This algorithm and its assumptions form the core of our discussion.

In section 2 we consider the connection of correlation and causation in general. Since a *single* correlation is not enough to infer a causal relation, we turn to the probabilistic and the causal structure of several variables in section 3. In section 4 we state the d-separation criterion and the stability assumption, which connect the causal to the probabilistic structure. In section 5 we review the inductive causation algorithm and, in section 6, discuss the assumptions underlying it. Finally, in section 7, we draw conclusions from our discussion.

2 Correlation and Causation

Correlation is perhaps the most frequently used concept in applied statistics. Its standard measure is the correlation coefficient, which assesses what can be called the intensity of linear relationship between two measures. Correlation is closely related to probabilistic dependence. However, the two concepts are not identical, because zero correlation does not imply independence. But since this difference is of no importance for our discussion, we use the term "correlation" in the vernacular sense, i.e., as a synonym for (probabilistic) dependence.

Note that neither in the narrower statistical nor in the wider vernacular sense correlation is connected directly to causal relation. We usually do not know *why* a correlation exists or does not exist, only *that* it is present or not. Nevertheless such erroneous interpretation is tempting [5]:

> Much of the fascination of statistics lies embedded in a gut feeling — and never trust a gut feeling — that abstract measures summarizing large tables of data must express something more real and fundamental than the data itself. (Much professional training in statistics involves a conscious effort to counteract this gut feeling.) The technique of *correlation* has been particularly subject to such misuse because it seems to provide a path for inferences about causality. [...] [But t]he inference of cause must come from somewhere else, not from the simple fact of correlation — though an unexpected correlation may lead us to search for causes so long as we remember that we may not find them. [...] The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning.

It is easily demonstrated that indeed the vast majority of all correlations are, without doubt, noncausal. Consider, for example, the distance between the continents America and Europe over the past twenty years (or any other suitable period). Due to continental drift this distance increases a few centimeters every year. Consider also the average price of Swiss cheese in the United States over the same period.[1] The correlation coefficient of these two measures is close to 1, i.e., even in the narrow statistical sense they are strongly correlated. But obviously there is no causal relation whatsoever between them.

[1] We do not know much about the average price of Swiss cheese in the United States over the past twenty years, but we assume that it has risen. If it has not, substitute the price of any other consumer good that has.

Of course, we could have used also a lot of other measures that increased over the past years, for example, the distance of Halley's comet (since its last visit in 1986) or the reader's age. The same can be achieved with measures that *de*creased over the past years. Therefore, causality may neither be inferred from correlation with certainty (since there are counterexamples), nor even inferred with a high probability (since causal correlations themselves are fairly rare).

According to these arguments it seems to be a futile effort to try to infer causation from observed dependences. Indeed, there is no way to causation from a *single* correlation (i.e., a dependence between two variables). But this does not exclude immediately the possibility to infer from a set of (conditional) dependences and independences between *several* variables something about the underlying causal influences. There could be connections between the causal and the probabilistic *structure*, which enable us to discover the former at least partly.

3 Probabilistic and Causal Structure

From the point of view of statistics the basic idea underlying Bayesian networks is that a probability distribution P on a multi-dimensional domain can, under certain conditions, be decomposed into a set $\{P_1, \ldots, P_n\}$ of (conditional) distributions on lower-dimensional subspaces. Such a decomposition rests on two things: the chain rule of probability and a set of (conditional) independence statements. Let $\mathcal{U} = \{X_1, \ldots, X_n\}$ be a set of discrete random variables. Then the *chain rule of probability* states that $\forall x_1 \in \mathrm{dom}(X_1)$, ..., $x_n \in \mathrm{dom}(X_n)$:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | x_1, \ldots, x_{i-1}).$$

where $P(x_1, \ldots, x_n)$ is short for $P(X_1 = x_1, \ldots, X_n = x_n)$, etc. If a set of conditional independence statements is given, this factorization can sometimes be significantly simplified. Thus one arrives at $\forall x_1 \in \mathrm{dom}(X_1)$, ..., $x_n \in \mathrm{dom}(X_n)$:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | \pi_{\mathrm{par}(X_i)}(x_1, \ldots, x_{i-1})),$$

where $\pi_S(I)$ denotes the *projection* of an instantiation I of a set of random variables to the variables in S and $\mathrm{par}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$ is chosen in such a way that $X_i \perp\!\!\!\perp \{X_1, \ldots, X_{i-1}\} \setminus \mathrm{par}(X_i) \mid \mathrm{par}(X_i)$, i.e., that $\forall x_1 \in \mathrm{dom}(X_1), \ldots, x_i \in \mathrm{dom}(X_i) : P(x_i | \pi_{\mathrm{par}(X_i)}(x_1, \ldots, x_{i-1})) = P(x_i | x_1, \ldots, x_{i-1})$.

Such a factorization is usually represented by a *directed acyclic hypergraph*, in which each node represents a random variable and each hyperedge represents a conditional probability distribution. We need *hyper*edges, which connect more than two nodes, since in general a variable is conditioned on more than one other variable. But since each node can have at most one hyperedge leading to it, one may also use a normal directed graph. In this case all parent nodes of a given node are in the condition part of the distribution for that node. This directed (hyper)graph we call a *probabilistic structure* $\mathcal{P}$. It is obvious that it is not unique, since it depends on the ordering of the variables.

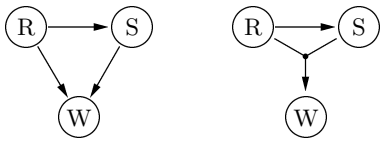


Fig. 1. Causal (left) and probabilistic structure (right) of the lawn example

We now turn to the causal structure of a domain. Our intuition of causation is perhaps best captured by a binary predicate "X (directly) causes Y" or "X has a (direct) causal influence on Y", where X is the *cause* and Y the *effect*. This predicate is usually seen as antisymmetric, i.e., if "X (directly) causes Y" holds, then "Y (directly) causes X" does not hold. Thus there is an inherent direction in causal influence, which seems to be a characteristic property of causation. For the most part it is due to our intuition that a cause precedes its effect in time.

Another formal interpretation is that an effect is as a function of its cause. But we reject this interpretation for several reasons. The first is that it brings in an assumption through the back door, which we want to make explicit (see section 6). Secondly, a function is not necessarily antisymmetric and thus cannot always represent the direction of causation. Thirdly, if one variable is a function of another, then there need not be a causal connection (see section 2). Hence functional dependence and causal influence should not be identified.

Because of the inherent direction, we can use a *directed hypergraph* to represent causal influences. (Where a *hyper*edge shows that a conjunction of causes is needed, and separate (normal) edges show, that each of several causes can lead to an effect. However, usually no harm is done, if a hyperedge is split into a set of normal edges.) This structure we call the *causal structure* $\mathcal{C}$. In principle directed loops, i.e., circular causal influences, are possible. (Such cycles are often exploited for control mechanisms, for example Watt's conical pendulum governor of the steam engine.) Nevertheless we do not consider circular causal structures, but assume that the causal influences form a *directed acyclic graph*.

A very simple and often used example is the following [15]: If it rains (R), the lawn will get wet (W). But it will also get wet, if the sprinkler (S) is turned on. In addition, if it rains, we will not turn on the sprinkler. Obviously, both R and S have a causal influence on W and R has a causal influence on S (though mediated through a human). These influences are represented by the causal structure shown on the left of figure 1. The probabilistic structure shown on the right in figure 1 (which, of course, is not unique) is very close to the causal structure, its corresponding normal graph would be identical to it.

4 *d*-Separation and Stability

It is obvious that storing both a *probabilistic structure* as well as a *causal structure* for a given domain is redundant. For instance, a cause and one of its direct effects should be dependent probabilistically. Thus the question arises, how the structures can be combined. The most promising approach seems to be to look for a method to read from the causal structure the independence statements that hold in the corresponding probabilistic structure. The best-known suggestion for

such a method is the so called d-separation criterion, of which it is claimed that it allows to determine whether two variables (or two sets of variables) are conditionally independent given a set S of variables: they are, if they are d-separated by S in the causal structure. d-separation is defined as follows [15, 4]:

Definition 1. *If R_1, R_2, and S are three disjoint subsets of nodes in a directed acyclic graph, then S is said to* d-separate *R_1 from R_2, iff there is no path from a node in R_1 to a node in R_2 along which the following two conditions hold:*

1. *every node with converging edges either is in S or has a descendant in S,*
2. *every other node is not in S.*

A path satisfying the conditions above is said to be active*; otherwise it is said to be* blocked *(by S). A* path *is a sequence of consecutive edges (of any direction).*

Note that the d-separation criterion does not say anything about the dependence or independence of R_1 and R_2 given S, if R_1 and R_2 are *not* d-separated by S. Usually this is sufficient, if a Bayesian network is to be constructed, since for applications it is not essential to find and represent *all* independences. However, we need more to infer causal structure. Therefore it is assumed that in a sampled probability distribution $\hat{P}$ there exist exactly those independences that can be read from the causal structure $\mathcal{C}$ using d-separation. This assumption is called *stability* [14] and can be formalized as ($R_1 \perp\!\!\!\perp R_2 \mid S$ holds in $\widehat{P}$) $\Leftrightarrow$ (S d-separates R_1 and R_2 in $\mathcal{C}$), where R_1, R_2, and S are sets of variables. Note that the stability assumption states that there is "no correlation without causation" (also known as Reichenbach's dictum), since between two variables that are dependent given any set of other variables, there must be a direct causal influence.

An important property of d-separation and the stability assumption is that they distinguish a common effect of two causes from the mediating variable in a causal chain and from the common cause of two effects. In the structures $A \rightarrow B \rightarrow C$ and $A \leftarrow B \rightarrow C$, A and C are independent given B, but in the structure $A \rightarrow B \leftarrow C$ they are not. This alleged asymmetry, studied earlier in [16], makes the inferences of the inductive causation algorithm [14] possible.

5 Inductive Causation

Even with the d-separation criterion and the stability assumption there are usually several causal structures that are compatible with the observed (conditional) dependences and independences. The main reason is that d-separation and stability cannot distinguish between causal chains and common causes. But in certain situations all compatible causal structures have a common substructure. The aim of the inductive causation algorithm is to find these invariant substructures.

The only ingredients of the inductive causation algorithm apart from the d-separation criterion and the stability assumption are the notions of a latent structure and of its projection. A *latent structure* is simply a causal structure in which some variables are unobservable (as it is often the case in real world problems). To handle such hidden variables, the notion of a projection of a latent structure is introduced. The idea is to restrict the number and influence of latent variables while preserving all dependences and independences.

Input: $\widehat{P}$, a sampled distribution over $\mathcal{U}$, the universe of discourse.
Output: core($\widehat{P}$), a marked hybrid acyclic graph.

1. For each pair of variables X and Y, search for a set $S_{XY} \subseteq \mathcal{U} \backslash \{X, Y\}$ such that $X \perp\!\!\!\perp Y \mid S_{XY}$ holds in $\widehat{P}$, i.e., X and Y are independent in $\widehat{P}$ conditioned on S_{XY}. If there is no such S_{XY}, place an undirected edge between the variables.
2. For each pair of non-adjacent variables X and Y with a common neighbour Z (i.e., Z is adjacent to X as well as to Y), check whether $Z \in S_{XY}$. If it is not, add arrowheads pointing to Z, i.e., $X \rightarrow Z \leftarrow Y$.
3. Form core($\widehat{P}$) by recursively adding arrowheads according to the following two rules:
   - If for two adjacent variables X and Y there is a strictly directed path from X to Y not including the edge from X to Y, then add an arrowhead pointing to Y.
   - If there are three variables X, Y, and Z with X and Y not adjacent, $Y - Z$, and either $X \rightarrow Z$ or $X \leftrightarrow Z$, then direct the link $Z \rightarrow Y$.
4. For each triplet of variables X, Y, and Z: If X and Y are not adjacent, $Z \rightarrow Y$, and either $X \rightarrow Z$ or $X \leftrightarrow Z$, then mark the edge $Z \rightarrow Y$.

Fig. 2. The Inductive Causation Algorithm [14]

Definition 2. *[14] A latent structure L_1 is a* projection *of another latent structure L_2, if and only if*

1. *Every unobservable variable in L_1 is a parentless common cause of exactly two non-adjacent (i.e., not directly connected) observable variables.*
2. *For every stable distribution P_2 which can be generated by L_2, there exists a stable distribution P_1 generated by L_1 such that $\forall X, Y \in O, S \subseteq O \backslash \{X, Y\}$: ($X \perp\!\!\!\perp Y \mid S$ holds in $P_2|_O$) $\Rightarrow$ ($X \perp\!\!\!\perp Y \mid S$ holds in $P_1|_O$), where O is the set of observable variables and $P|_O$ denotes the marginal probability distribution on these variables.*

(A stable distribution satisfies the stability assumption, i.e., exhibits only those independences identifiable by the d-separation criterion.)

It can be shown that for every latent structure there is at least one projection. Note that a projection must exhibit only the same (in)dependence structure (w.r.t. *d*-separation), but need not be able to generate the same distribution.[2] In essence, the notion of a projection is only a technical trick to be able to represent dependences that are due to latent variables by bidirected edges (which are an intuitive representation of a hidden common cause of exactly two variables).

One thus arrives at the inductive causation algorithm [14] shown in figure 2. Step 1 determines the variable pairs between which there must exist a direct causal influence or a hidden common cause, because an indirect influence should

[2] Otherwise a counterexample could easily be found: Consider seven binary variables A, B, C, D, E, F, and G, i.e., $\mathrm{dom}(A) = \mathrm{dom}(B) = \ldots = \mathrm{dom}(G) = \{0, 1\}$. Let A be hidden and $E = A \cdot B$, $F = A \cdot C$, and $G = A \cdot D$. A projection of this structure contains three latent variables connecting E and F, E and G, and F and G, respectively. It is easy to prove that such a structure cannot generate a stable probability distribution that can be generated by the original structure.

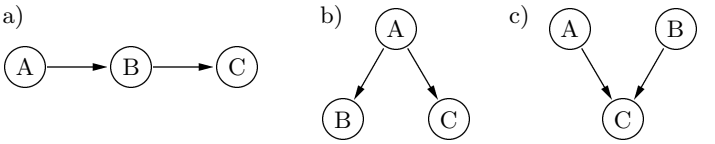


Fig. 3. Possible causal connections

enable us to find a set S that renders the two variables independent. In step 2 the asymmetry inherent in the d-separation criterion is exploited to direct edges towards a common effect. Part 1 of step 3 ensures that the resulting structure is acyclic. Part 2 uses the fact that $Y \rightarrow Z$ is impossible, since otherwise step 2 would have already directed the edge in this way. Finally, step 4 marks those unidirected links that cannot be replaced by a hidden common cause (based on similar grounds as part 2 of step 3). The output core has four kinds of edges:

1. marked unidirected edges representing *genuine causal influences* (which must be direct causal influences in a projection),
2. unmarked unidirected edges representing *potential causal influences* (which may be direct causal influences or brought about by a hidden common cause),
3. bidirected edges representing *spurious associations* (which are due to a hidden common cause in a projection), and
4. undirected edges representing unclassifiable relations.

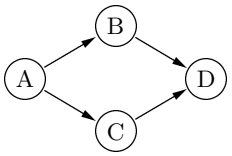
6 Critique of the Underlying Assumptions

In this section we discuss the assumptions underlying d-separation and stability by considering some special cases with only few variables. The simplest case are causal chains, like the one shown in figure 3.a. If a variable has a direct causal influence on another, they should be dependent at least unconditionally, i.e., $A \not\perp\!\!\!\perp B \mid \emptyset$ and $B \not\perp\!\!\!\perp C \mid \emptyset$. It is also obvious, that $A \perp\!\!\!\perp C \mid \{B\}$. A direct cause, if fixed, should shield the effect from any change in an indirect cause, since a change in the indirect cause can influence the effect only by changing the direct cause. But to decide whether B and C are dependent given A or not, we need to know the causal influences in more detail. For instance, if $B = f(A)$ and $C = g(B)$, then $B \perp\!\!\!\perp C \mid A$. But if the value of A does not completely determine the value of B (just as the rain did not completely determine the state of the sprinkler in the lawn example), then B and C will usually be dependent. Although the former is not uncommon, the stability assumption excludes it.

The next cases are diverging or converging causal influences, like those shown in figures 3.b and 3.c. The main problems with these structures are whether $B \perp\!\!\!\perp C \mid \{A\}$ (in 3.b) and $A \perp\!\!\!\perp B \mid \{C\}$ (in 3.c) hold or not. The assumptions by which d-separation and the stability assumption handle this difficulty are:

Common Cause Assumption (Causal Markov Assumption).
Given *all* of their (direct or indirect) common causes, two effects are independent, i.e., in figure 3.b the variables B and C are independent given A. If B and C are still dependent given A, it is postulated that either B has a causal influence on C or vice versa or there is another (hidden) common cause of B and C (apart from A). That is, the causal structure is considered to be incomplete.

Fig. 4. Interaction of common cause and common effect assumption



Common Effect Assumption.
Given *one* of their (direct or indirect) common effects, two causes are dependent, i.e., in figure 3.c the variables A and B are dependent given C. For applications, this assumption is less important than the previous one, since nothing is lost, if it is assumed that A and B are dependent given C though they are not. Only the storage savings resulting from a possible decomposition cannot be exploited.

Note that the common cause assumption necessarily holds, if causation is interpreted as functional dependence. Then it only says that fixing all the arguments that (directly or indirectly) enter both functions associated with the two effects renders the effects independent. But this is obvious, since any variation still possible has to be due to independent arguments that enter only one function. This is the main reason why we rejected this interpretation of causation. It is not at all obvious that causation should satisfy the common cause assumption.

A situation with diverging causal influences also poses another problem: Are B and C independent unconditionally? In most situation they are not, but if, for example, $\mathrm{dom}(A) = \{0, 1, 2, 3\}$, $\mathrm{dom}(B) = \mathrm{dom}(C) = \{0, 1\}$ and $B = A \bmod 2$, $C = A \operatorname{div} 2$, then they will be. The stability assumption rules out this possibility.

The two assumptions also interact and this can lead to a priority problem. For example in figure 4: Given A as well as D, are B and C independent? The common cause assumption affirms this, the common effect assumption denies it. Since the stability assumption requires B and C to be dependent, it contains the assumption that in case of a tie the common effect assumption has the upper hand. Note that from strict functional dependence $B \perp\!\!\!\perp C \mid \{A, D\}$ follows.

In the following we examine some of the assumption identified above in more detail, especially the common cause and the common effect assumption.

Common Cause Assumption (Causal Markov Assumption)
Consider an arrangement of tubes like the one shown in figure 5.a. If a ball is dropped into this arrangement, it will reappear at one of the two outlets. If we neglect the time it takes the ball to travel through the tubes, we can define three binary variables T, L, and R indicating whether there is a ball at the top T, at the left outlet L or at the right outlet R. Obviously, whether there is a ball at T or not has a causal influence on L and on R. But L and R are dependent given T, because the ball can reappear only at one outlet.

At first sight the common cause assumption seems to fail in this situation. However, we can always assume that there is a hidden common cause, for instance, an imperfectness of the ball or the tubes. If we knew the state of this cause, the outlet at which the ball will reappear could be determined and hence the common cause assumption would hold. Obviously, if there is a dependence between two effects, we can always say that there must be another hidden common cause. We just did not find it, because we did not look hard enough. Since this is a statement of existence, it cannot be disproven.

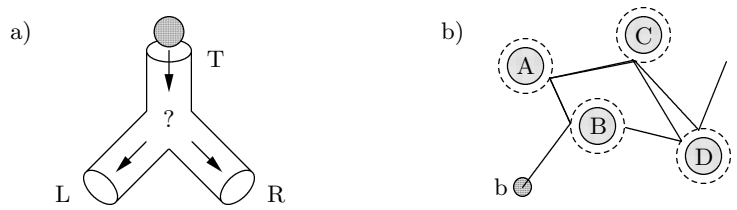


Fig. 5. a) Y-shaped tube arrangement into which a ball is dropped. Since it can reappear only at L *or* at R, but not at both, the corresponding variables are dependent. b) Billiard with round obstacles exhibits sensitive dependence on the initial conditions.

The idea that, in principle, we could discover the causes that determine the way the ball goes is deeply rooted in the mechanistic paradigm of physics, which is perhaps best symbolized by Laplace's demon.[3] But quantum theory suggests that such a view is wrong [1, 13]: It may very well be that even if we look hard enough, we will not find a hidden common cause to explain the dependence.

To elaborate a little: Among the basic statements of quantum mechanics are Heisenberg's uncertainty relations. One of these states that $\triangle x \cdot \triangle p_x \geq \frac{\hbar}{2}$. That is, we cannot measure both the location x and the momentum p_x of a particle with arbitrary precision in such a way that we can predict its exact trajectory. There is a finite upper bound due to the unavoidable interaction with the observed particle. However, in our example we may need to predict the exact trajectory of the ball in order to determine the outlet with certainty.

The objection may be raised that $\frac{\hbar}{2}$ is too small to have any observable influence. To refute this, we could add to our example an "uncertainty amplifier" based on the ideas studied in chaos theory, i.e., a system that exhibits a sensitive dependence on the initial conditions. A simple example is billiard with round obstacles [17], as shown in figure 5.b. The two trajectories of the billiard ball b, which in the beginning differ only by about $\frac{1}{100}$ degree, differ by about 100 degrees after only four collisions. (This is a precisely computed example, not a sketch.) Therefore, if we add a wider tube containing spheres or semi-spheres in front of the inlet T, it is plausible that even a tiny change of the position or the momentum of the ball at the new inlet may change the outlet at which the ball will reappear. Therefore quantum mechanical uncertainty cannot be neglected.

Another objection is that there could be "hidden parameters", which, if discovered, would remove the statistical nature of quantum mechanics. However, as [13] showed[4], this is tantamount to claiming that quantum mechanics is false — a claim for which we do not have any convincing evidence.

[3] Laplace wrote [7]: "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at any given moment knew all the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom: for such an intellect nothing would be uncertain; and the future just like the past would be present before its eyes."

[4] v. Neumann wrote: "[...] the established results of quantum mechanics can never be re-derived with their [the hidden parameters'] help. In fact, we have even ascertained that it is impossible that the same physical quantities exist with the same function

murderer	death	other	$\sum$
black	59	2448	2507
white	72	2185	2257
$\sum$	131	4633	4764

Table 1. Death sentencing and race in Florida 1973–1979. The hypothesis that the two variables are independent can be rejected only with an error probability greater than 7.8% (according to a χ^2 test).

victim	murderer	death	other
black	black	11	2209
	white	0	111
white	black	48	239
	white	72	2074

Table 2. Death sentencing and race in Florida 1973–1979, full table. For white victims the hypothesis that the two other variables are independent can be rejected with an error probability less than 0.01% (according to a χ^2 test).

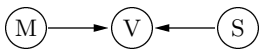


Fig. 6. Core inferred by the inductive causation algorithm for the above data.

Common Effect Assumption

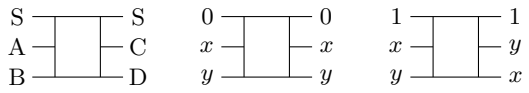
According to Salmon [18], it seems to be hard to come up with an example in which the common effect assumption does not hold. Part of the problem seems to be that most macroscopic phenomena are described by continuous real-valued functions, but there is no continuous n-ary function, $n \geq 2$, which is injective (and would be a simple, though not the only possible counterexample).

However, there are real world examples that come close, for instance, statistical data concerning death sentencing and race in Florida 1973–1979 (according to [8] as cited in [22]). From table 1 it is plausible to assume that *murderer* and *sentence* are independent. Splitting the data w.r.t. *victim* shows that they are strongly dependent given this variable (see table 2). Hence the inductive causation algorithm yields the causal structure shown in figure 6. But this is not acceptable: A direct causal influence of *sentence* on *victim* is obviously impossible (since the sentence follows the murder in time), while a common cause is hardly imaginable. The most natural explanation of the data, namely that *victim* has a causal influence on *sentence*, is explicitly ruled out by the algorithm.

This example shows that an argument mentioned in [14] in favour of the stability assumption is not convincing. It refers to [20], where it is shown that, if the parameters of a distribution are chosen at random from any reasonable distribution, then any unstable distribution has measure zero. But the problem is that this is not the correct set of distributions to look at. When trying to infer causal influence, we have to take into account all distributions that *could be mistaken for an unstable distribution*. Indeed, the true probability distribution in our example may very well be stable, i.e., *murderer* and *sentence* may actually

---

connections [...], if other variables (i.e., "hidden parameters") should exist in addition to the wave functions. Nor would it help if there existed other, as yet undiscovered physical quantities, [...], because the relations assumed by quantum mechanics [...] would have to fail already for the known quantities [...] It is therefore not, as often assumed, a question of a re-interpretation of quantum mechanics, — the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible."



S	0	0	0	0	1	1	1	1
A	0	0	1	1	0	0	1	1
B	0	1	0	1	0	1	0	1
C	0	0	1	1	0	1	0	1
D	0	1	0	1	0	0	1	1

Fig. 7. The Fredkin gate [2]

be marginally dependent. But the distribution in the sample is so close to an independent distribution that it may very well be confused with one.

In addition, the special parameter assignments leading to unstable distributions may have high probability. For example, it would be reasonable to assume that two variables are governed by the same probability distribution, if they were the results of structurally equivalent processes. Yet such an assumption can lead to an unstable distribution, especially in a situation, in which common cause and common effect assumption interact. For instance, for a Fredkin gate [2] (a universal gate for computations in conservative logic, see figure 7), the two outputs C and D are independent, if the two inputs A and B assume value 1 with the same probability. In this case, as one can easily verify, the causal direction assigned to the connection A—C depends on whether the variables A, B, and C or the variables A, C, and D are observed.

7 Conclusions

The discussion of the assumptions underlying the inductive causation algorithm showed that at least some of them can be reasonably doubted. In addition, the inductive causation algorithm cannot deal adequately with accidental correlations. But we saw in section 2 that we sometimes reject a causal explanation in spite of the statistical data supporting such a claim. In our opinion it is very important for an adequate theory of causation to explain such a rejection.[5] In summary, when planning to apply this algorithm, one should carefully check whether the assumptions can be accepted and whether the underlying interpretation of causality is adequate for the problem at hand.

A related question is: Given a causal relation between two variables, we are usually much more confident in an inference from the state of one of them to the state of the other than we would be, if our reasoning was based only on a number of similar cases we observed in the past. But the inductive causation algorithm infers causation from a set of past observations, namely a sampled probability distribution. If the result is not substantiated by other means, can we be any more confident in our reasoning than we would be, if we based it directly on the observed correlations? It seems to be obvious that we can not. Hence the question arises whether the inductive causation algorithm is more than just a heuristic method to point out possible causal connections, which than have to be further investigated. Of course, this does not discredit the inductive causation algorithm, since good heuristics are a valuable thing to have.

[5] An approach to causation that does not suffer from this deficiency was suggested by Lorenz and later developed e.g. in [21]. It models causal connections as a transfer of energy. [12] suggests a closely related model.

References

1. R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics 3: Quantum Mechanics.* Addison Wesley, Reading, MA, USA 1965
2. E. Fredkin and T. Toffoli. Conservative Logic. *Int. J. Theoretical Physics*, 21(3/4):219–253, 1982
3. J. Gebhardt and R. Kruse. POSSINFER — A Software Tool for Possibilistic Inference. In: D. Dubois, H. Prade, and R. Yager, eds. *Fuzzy Set Methods in Information Engineering: A Guided Tour of Applications*, J. Wiley & Sons, Chichester, England 1995
4. D. Geiger, T.S. Verma, and J. Pearl. Identifying Independence in Bayesian Networks. *Networks* 20:507–534, J. Wiley & Sons, Sussex, England 1990
5. S.J. Gould. *The Mismeasure of Man.* Penguin Books, New York, NY, USA 1992
6. D. Hume. An Enquiry Concerning Human Understanding. First published in: *Essays and Treatises on Several Subjects*, 1777. Reprinted in: C. Brighton and P. Edwards, eds. *David Hume on Human Nature and the Understanding.* Collier Classics in the History of Thought, Macmillan, New York, NY, USA 1962
7. M. Kline. *Mathematics — The Loss of Certainty.* Oxford University Press, New York, NY, USA 1980
8. K. Krippendorf. *Information Theory and Statistics.* Quantitative Applications in the Social Sciences 62. Sage Publications, London, England 1986
9. R. Kruse, E. Schwecke, and J. Heinsohn. *Uncertainty and Vagueness in Knowledge-based Systems: Numerical Methods.* Springer, Berlin, Germany 1991
10. R. Kruse, J. Gebhardt, and F. Klawonn. *Foundations of Fuzzy Systems*, J. Wiley & Sons, Chichester, England 1994
11. S.L. Lauritzen and D.J. Spiegelhalter. Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *J. of the Royal Statistical Society, Series B*, 2(50):157–224, 1988
12. J.F. Lemmer. The Causal Markov Condition, Fact or Artifact? *SIGART Bulletin*, 7(3):3-16, 1996
13. J. von Neumann. *Mathematical Foundations of Quantum Mechanics (12th edition).* Princeton University Press, Princeton, NJ, USA 1996
14. J. Pearl and T.S. Verma. A Theory of Inferred Causation. *Proc. 2nd Int. Conf. on Principles of Knowledge Representation and Reasoning*, Morgan Kaufman, San Mateo, CA, USA 1991
15. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd edition).* Morgan Kaufman, San Mateo, CA, 1992
16. H. Reichenbach. *The Direction of Time.* University of California Press, Berkeley, CA, USA 1956
17. D. Ruelle. *Zufall und Chaos.* Springer, Heidelberg, Germany 1993
18. W.C. Salmon. *Scientific Explanation and the Causal Structure of the World.* Princeton University Press, Princeton, NJ, USA 1984
19. G. Shafer and P.P. Shenoy. *Local Computations in Hypertrees.* WP 201, School of Business, University of Kansas, Lawrence, KA, USA 1988
20. P. Spirtes, C. Glymour, and R. Scheines. *Causality from Probability.* TR CMU-LCL-89-4, Carnegie Mellon University, Pittsburgh, PA, USA 1989
21. G. Vollmer. Ein neuer dogmatischer Schlummer? — Kausalität trotz Hume und Kant. In: *Akten des 5. Int. Kant-Kongresses Mainz 1981*, 1125–1138, Bouvier, Bonn, Germany 1981
22. J. Whittaker. *Graphical Models in Applied Multivariate Statistics.* J. Wiley & Sons, Chichester, England 1990