# Notes on Causality, Prediction and Search by Peter Spirtes

D. Gueorguiev, 9/13/23

## Notation and Basic Definitions:

### Graphs:

Notation for Edges:

**undirected edge** :

**directed edge** :

**non-directed edge** :

**partially directed edge** :

Note on Inducing Path Graph: contains both directed edges (e.g., ) , bi-directed edges (e.g., ), non-directed edges (e.g., ), and partially directed edges (e.g., ).

**Graph** (*traditional definition*): ordered pair where is a set of vertices and is a set of edges. The members of are pairs of vertices (an ordered pair in a directed graph and an unordered pair in an undirected graph). For example, the edge is represented by the ordered pair . We need to specify variables and **marks** at each end. In general, we will allow that the end of an edge can be unmarked, can be marked with an arrowhead , or can be marked with an . For example, the left end of can be represented as ordered pair , while the right end can be represented as the ordered pair . The entire edge is a set of ordered pairs representing the endpoints . The edge is the same as .

Note that a directed edge such as has no mark at the endpoint; we consider the mark at the A endpoint to be empty, but when we write out the ordered pair, we will use the notation to stand for the empty mark e.g.,

**Graph** (our definition): an ordered triple where is a non-empty set of vertices, is a non-empty set of marks, and is a set of sets of ordered pairs of the form , where and are in , , and and are in . Except in our discussion of systems with feedback we will always assume that in any graph, any pair of vertices and occur in at most one set in , or, in other words, that there is at most one edge between any two vertices. If we say that is over .

Figure 1: Example of directed graph

For example, the directed graph on Figure 1 can be represented as:

**edge**: any member of .

**edge-end**: each ordered pair

**endpoint**: each vertex in an edge

**adjacent endpoints:** vertices are adjacent iff there is an edge with endpoints

**undirected graph**: a graph in which the set of marks .

**directed graph**: a graph in which the set of marks .

**directed edge** from to : an edge

**edge into** : any edge

**edge out** of **:** any edge

**parent/child**: is parent of which is child of if there is a directed edge from to

**indegree** of vertex : the number of the parents of

**outdegree** of vertex : the number of children of

**undirected path** between and in graph : a sequence of vertices beginning with and ending with such that for every pair of vertices and that are adjacent in the sequence there is an edge in .

**edge is in path**: is in the path iff and are adjacent to each other (in either order) in .

vertices **adjacent on path**: if an edge between and is in the path we say that and are adjacent on .

**path is out of vertex**: if the edge containing in **an undirected** **path** between and **is** **out of** then we say that **path is out** of .

**path is into vertex**: if the edge containing in a path between and is into we say that **the path is into** .

**empty path**: sequence which consists of a single vertex.

**acyclic path**: a path which contains no vertex more than once; otherwise, it is **cyclic path**.

: For a given graph and vertex , is the set of parents of .

: For a given graph and vertex , is the set of descendants of .

//TODO: finish the section on Graph basic definitions

### Probability

We will use sometimes the special summation symbol which has the following properties:

1. when sets of random variables are written beneath the special summation symbol, it is understood that the summation is to be taken over sets of values of the random variables, not the random variables themselves
2. if a conditional probability distribution appears in the scope of such summation symbol, the summation is to be taken only over values of the random variables for which the conditional probability distributions are defined.
3. If there are no values of the random variables under the special summation symbol for which the conditional probability distributions in the scope of the symbol are defined, then the summation is equal to zero.

The vertices of the graphs we consider will always be random variables taking values in one of the following: the real line, the nonnegative reals, an interval of integers

By a joint distribution on the vertices of a graph we mean a countably additive probability measure on the Cartesian product of these objects. We say that two random variables and are **independent** when the joint density of is the product of the density of and the density of for all values of and . We write this as . We say that a set of random variables is **jointly independent** when any two disjoint subsets of the set are independent of one another. We say that random variables are independently conditional on (or given ) when the density of , given equals the product of the density of given and the density of given , for all values of and for all values of for which the density of is not equal to 0. Generalization for set of random variables, : if is independent of given we write and we say that the **order of the conditional independence** is equal to the number of variables in .

In the discrete case, we say that a distribution over is positive iff for all values of , .

If is included in and

// Note to myself: this is a very obscure way to write =

we say that is the **marginal** of over .

### Graphs and Probability Distributions

We are interested in graphical representation of conditional independence relations in a distribution

**Definition**: *Markov condition*:

A directed acyclic graph over and a probability distribution P(V) satisfies the *Markov condition* if and only if for every in , is independent of given . In the terminology of (Pearl, 1988) is an **I-map** of .

//TODO: finish the basic definitions in Graphs and Probability Distributions

//TODO: finish basic definitions

## Introductory Notes

### Notes on Chapter 1.1 The Issue

Claim is made that existing methods for function fitting through some minimization technique *cannot* be reliably used to infer causal dependencies from random uncontrolled samples or predict the effects of manipulation. Linear regression is taught as means of fitting a line to sample data and also as means of predicting new values of a variable. Many of the real applications of regression are to predict values of a variable when the regressors are *manipulated*, that is, when an action or policy forces some new distribution to the regressors. Lots of claims are circulated that regression cannot be used to infer causes from non-experimental samples, or to make predictions about manipulated systems. Bolder claim is made: it is not possible for any statistical method to be used to reliably infer causes or to make prediction about manipulated systems.

*The Arguments against the possibility of reliable inference of causal structure from statistical samples*

* Many examples in which regression yields erroneous causal inferences
* Underdetermination of causal structure by statistical dependencies

Claim often seen in the literature of at the time: Statistical dependency is possible to be observed between sample values of variables X and Y for the following:

1. When causes
2. When causes
3. When each causes the other
4. When some third variable causes both
5. When the sample is not representative
6. When the values of and form time series

However, immediate warning is issued that “experimentation, and not the existence of statistical relationship” is necessary to establish a cause.

Counterclaim by the authors: It is true enough that we cannot distinguish among (i), (ii), (iii) and (iv) when we measure only two variables, X and Y, but what is the proof that we cannot distinguish among these alternative causal relations if additional variables are measured? We know that in other contexts identifiability and estimation properties of parameters relation a pair of variables can be changed if further variables are measured; the method of instrumental variables is an illustration. May the same not be the case with identifying causal structure?

Consider (v) : this claim should not be taken seriously as anything in which statistics is involved would be dismissed if the sample is notchosen well, that is – it is not representative. (vi) makes a sound point.

Note on *collapsibility*:

Important theoretical question: when can the same conclusions about the existence and strength of an influence of one variable, , on another, , be obtained by analyzing a reduced set of variables, (containing and ), rather than a larger set of variables that properly includes ? When and how can the analysis of a set of variables, including and , reliably determine whether variable causes , even though there may be common causes of and that are not in ; When and how, in such circumstances, can we reliably predict the effect on of manipulating ? Another related question – when are the log-linear parameters for a model obtained by marginalizing out some variables the same as the corresponding parameters of the larger, unmarginalized model?

Note on *model selection* or *specification search*:

“Model selection” or “specification search” is an area of research in which questions of causal inference have been obscured. Statistical models such as log-linear models, structural equations models, regression models, etc have two distinct roles. One role is to restrict the class of possible probability distributions among a set of variables and to parametrize the family of distributions that satisfy the restriction. Thus, a log-linear model, for example, is given by specifying that particular parameters vanish in a linear expansion of the logarithm of the probability of any cell.

The importance of hypothesis selection in this respect is that the restrictions and the parametrization should aid one in understanding and efficiently estimating the distribution. The other role such models may have is to inform prediction – the predictions are often about the effects of actions or events that, if they were to occur, would *alter* the probability distribution. These are causal claims – they do not follow from any estimate of an actual probability distribution, and they depend on a further interpretation of the representations through which the restrictions of a statistical model are expressed. Statistical hypotheses used with a causal interpretation would seem either to be correct and the difference is important.

*Discussion on the lack of statistical theory for causal inference*

Question: Why so little statistical theory is concerned with causal inference? Some attempts were made to give accounts of causation entirely in terms of probability relations, while others try to characterize causation in terms of counterfactual conditions. There cannot be any rigorous theory about what is undefined.

However, the absence of a “definition” of causation does not prevent research which promotes causal inference in experimental contexts. The fact that one variable causes another cannot always depend essentially on whether we discover the fact by experimentation. Every interpretation of probability appeals to obscure counterfactual assumptions or to mysterious properties in a sense which is similar to the arguments regarding the definition of causality.

At the time of the writing of this work there are two attempts to express causality via stochastic formalisms, each valuable and neither sufficient by itself, hence the value of the current work.

A mathematical representation of causal dependencies among a set of variables has been in the statistical literature for a more than century. (Wright, 1934) used directed graphs to represent causal structures.

//TODO: finish this paragraph – p26 of the pdf

(Harri Kiiveri, 1982) related //TODO: finish this paragraph

//TODO: finish the notes on Chapter 1.1 The Issue

Notes on Chapter 1.3 Themes

*Fundamental questions in this thesis*: understanding the systematic connections between causal dependency and stochastic dependency. What are the limits to reliable causal inference given the presence of such connections? Under what conditions the causal inference will produce reliable results? What are the limitations for causal inference given our theoretical understanding between causality and probability? Within these limits we can investigate rigorously the shortcomings of the algorithms that search for causal structures from statistical properties of the datasets.

The asymptotic reliabilities of well-defined procedures can be determined mathematically, while short run behavior can be estimated/bounded through well-chosen simulation experiments. This thesis investigates into the theory of model specification search adopting results from the theory of estimation.

*Additional investigations*:

Understanding latent variables and their impact on causality.

This work investigates various asymptotic methods and their reliability to obtain information about the presence or absence of unmeasured common causes, and about their causal relations. *The Markov Condition* and *The Faithfulness Conditions* are widely assumed throughout this work. Informative sufficient conditions for the presence of unmeasured common causes are investigated under the assumption of Markov and Faithfulness conditions. Theorems about causal conclusions and predictions will be drawn whether or not latent variables are present. Tetrad Representation Theorem is more powerful theorem valid under the assumption of linearity – this theorem can be used to identify the presence of unmeasured common causes.

Conditions sufficient and/or necessary to correctly predict the effects of a policy applied to a population that has been studied through a sample, and whose causal structure is not known a priori.

*Related question*: when will the conditional distribution of on when is *forced* to have a certain value, equal the probability of conditional on , in an observational or experimental population?

One can find an answer assuming that certain counterfactual claims are known. That is, knowing aspects of the causal structure of the system under study.

*Follow-up question*: when can the relevant causal knowledge necessary to answer the question above be obtained from the sample data?

In this work results are described that answer the follow-up question and give information from sample data about when the conditional probability of on and is invariant under a manipulation of .

*Fundamental issue about prediction*:

Let us assume that the distribution of is to be directly manipulated. Let us denote the manipulated distribution with . *Question*: when can the resulting distribution of a set of variables conditional on a set be calculated from the distributions , of , , and other variables in an observational or experimental population in which were not manipulated for each unit in the sample?

In order to answer this question, we need to find out more about the formal connections between probability and causal structure.

*Another topic of discussion*: the commonly used statistical search procedures are sub-optimal for causal inference.

Automated model search procedures (especially linear and logistic regression) are asymptotically unreliable against alternative causal hypotheses that are often consistent with prior knowledge. If the procedures give the right answers in the ideal case of perfect information about the population distribution, one can look around for better tests and more computationally efficient algorithms. But if, as in the case of regression and many other automated techniques, probability relations and causal relations are incorrectly matched there is no way to produce good inferences and there is no point to investigate more performant algorithms based on flawed causal models.

*Another topic discussed in this work*: the importance of causal reasoning in the design of empirical studies.

The adopted view by many statisticians at the time is that causation is not something which can be established by data analysis. Establishing causation requires logical arguments that go beyond the realm of numerical manipulation. Establishing causation requires intrinsic understanding of the studied phenomenon beyond the set of experiments used in the analysis. When observational studies are used as a basis for causal inference the jump from correlation to causation must be made on nonstatistical grounds. The authors of this thesis question this view and argue that once a formal understanding of the connection between causal structure and probability is in place, the questions about the comparative power of experiment versus observation can be answered by a mathematical model based on the causal information that can be extracted from experimental and observational design.

//TODO: finish the paragraph on Chapter 1.3 Themes

# Bibliography

Eells, E. (1991). *Probabilistic Causality.* Cambridge UK: Cambridge University Press.

G.E. Hughes, M. C. (1996). *A New Introduction To Modal Logic.* London: Routledge.

Good, I. J. (1961). A Causal Calculus (I and II). *The British Journal for the Philosophy of Science*, 305-318, 43-51.

Harri Kiiveri, T. S. (1982). Structural Analysis of Multivariate Data: A Review. *Sociological Methodology*, pp. 209-289.

Hume, D. (1748). *Philosophical Essays Concerning Human Understanding.* London: Printed for A. Millar, opposite Katharine-Street in the Strand. MDCCXLVII.

Lewis, D. (1973). *Counterfactuals.* Malden, Massachusetts: Blackwell Publishers.

Lewis, D. (1974). Causation. *Journal of Philosophy*, 556-567.

Mackie, J. (1980). *The Cement of The Universe: A Study of Causation.* New York: Oxford University Press, New York, United States.

Otte, R. E. (1982). *Probability and Causality, PhD Thesis.* Ann Arbor , MI, 48106: University of Arizona Graduate College, University Microfilm International.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann.

Reichenbach, H. (1956). *The Direction of Time.* Berkeley and Los Angeles, California: University of California Press.

Salmon, W. C. (1980). Probabilistic Causality. *Pacific Philosophical Quarterly*, pp. 137-153.

Salmon, W. C. (1998). *Causality and Explanation.* Pittsburgh, Pennsylvania: Oxford University Press.

Spirtes, P., Glymour, C., & Sheines, R. (1993). *Causation, Prediction and Search.* New York: Springer Verlag.

Wright, S. (1934, September). The Method of Path Coefficients. *Annals of Mathematical Statistics, 5*(3), 00161-215.

# Downloadable Links for the Bibliography

(Eells, 1991): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/eells_probabilistic_causality_1991.pdf)

(Reichenbach, 1956): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/the-direction-of-time-hans-reichenbach-ucal-press-1971.pdf)

(Spirtes, Glymour, & Sheines, 1993): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/CausationPredictionandSearch_Spirtes_CMU_2000.pdf)

(Otte, 1982): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/Probability_and_Causality_PhD_Thesis_Otte_1982.pdf)

(G.E. Hughes, 1996): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/ModelChecking/huges_cresswell_modal_logic.pdf)

(Salmon, Causality and Explanation, 1998): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/Causality_and_Explanation_Wesley_Salmon_1997.pdf)

(Salmon, Probabilistic Causality, 1980): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/Probabilistic_Causality_Salmon_1980.pdf)

(Good, 1961): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/CausalCalculus_part_I_and_II_Good_1960.pdf)

(Mackie, 1980): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/cement-of-the-universe-a-study-of-causation-JL-Mackie-1980.pdf)

(Lewis, Counterfactuals, 1973): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/Counterfactuals-lewis-1973.pdf)

(Lewis, Causation, 1974): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/Lewis-Causation_1974.pdf)

(Pearl, 1988): [here](https://github.com/dimitarpg13/root_cause_analysis_and_model_checking/blob/main/literature/books/JudeaPearl-ProbabilisticReasoninginIntelligentSystems_1988.pdf)