# Causality, Probability and Time Notes

Book by Samantha Kleinberg

## Notes on Types of Causes and their Representation

### Insignificant Causes – Intuitions

The conditions for *prima facie* causality are insufficient to distinguish between causes and non-causes, and the primary difference between probabilistic theories of causality is in how exactly they make this distinction. The two main types of methods ar those based on information and those based on manipulation. The information-based theories use the idea that a cause provides some information about an effect that cannot be gained in other ways and set about finding evidence for that. Manipulation theories hold that a cause is a way of bringing about an effect and can be understood in terms of how the probability or value of the effect changes when manipulating the cause to be true. One approach is not inherently superior to the other – there are counterexamples to both manipulation and information-based methods and neither subsumes the other. Manipulation is not possible in many cases, and therefore it is undesirable to require it. On other side, methods that aim to infer causal relationships from observational data – because it is readily available or because manipulations are costly, unfeasible, or unethical – generally use some variant of the information-based approach. The distinguishing feature between all variants of the information-based methods is how to quantify the information provided about the effect. The basic idea is that of holding fixed some set of information and then see how likely the effect is when the cause is present and absent relative to that set of information.

Desiderata for quantifying causal influence

Consider the following set of relationships (an abstraction of DTMC):

This is a subset of a full system that we may have a model of or observe. Here can cause in two time units through two paths, directly or through d: of the time causes d at , of the time it causes directly at t+2, and  of the time it does neither. Now assuming the marginal probability of is much lower than how can we determine how significant and are for ?

One approach is to find the earliest cause that best accounts for the effect. This is Suppes’ method which says that a cause is spurious if there is other information that predicts the effect at least as well and, once this information is known, the probability of the effect is unchanged by knowledge of the spurious cause. In this example this method would determine erroneously that is a spurious cause since occurs earlier than and .

Similarly, using Granger causality one would also fail to find as a cause of since the probability of is unchanged once the information about is included. This is incorrect since d is in fact a cause of e and accounts for e exactly as well as c does

## Appendix

### Hume’s Regularity Theory

Hume’s interpretation of Causality is that the latter is routine occurrence of an effect following a cause. This is known as the Regularity Theory of Causality. There are many examples of *noncausal regularities* (*Example*: the presence of umbrella vendors and rain) and *causation without regularities* (*Example*: a drug causing death to a single patient). One of the major omissions in Hume’s Regularity Theory is that there is no method for separating which parts of a regular sequence of events are essential from those that are correlated with others.

*Example:* One may always open the door to my office, turn on the light and then begin writing papers or code on his computer. While the light is being useful for writing, it is not essential.

Many effects have multiple causes that are comprised of interacting components, such as the impact of environmental factors and genetic mutations on health. Hume’s approach does allow for reasoning with these types of causal complexes.

### Granger causality

#### Notation:

denotes the value of the variable at time

denotes the set of measurements of up to time i.e.,

denotes the set of all possible knowledge up to time (including both and )

**Definition A.1** Granger-causes if

(A.1)

As this is not exactly causality, it has come to be called Granger causality. Granger’s definition takes temporal priority as a given and does not make claims about how much of a difference makes to the probability of (or whether this difference is positive or negative). may not be the best or only predictor of , rather it is simply found to be informative after accounting for other information.

Granger’s test is understood usually as a *bivariate test*. In the bivariate test only two time series are included: that of the effect, and that of the cause, . One bivariate method is to use autoregressive model with the two variables where if the coefficients of the lagged values of are non-zero, then is said to Granger-cause

An -lag autoregressive model takes lagged values of time series up to when calculating the value of a variable at time . Each lagged value is weighted by a coefficient, so that a variable may depend more strongly on recent events than those that are more temporarily distant. More formally, can be represented as the following -lag linear autoregressive model:

(A.2)

The lags mean that the values of and at times influence that of at time . The coefficient indicates how much the value at depends on . Here means the influence of on , so that models the dependence of on itself and models that of on . Here can be finite or infinite – the latter was done in the original Granger work. The error term is assumed to be random variable with mean zero. Using this, Granger causality can be tested by whether non-zero values for lead to a smaller variance in the error term than when these are zero (and whether this reduction is statistically significant).

Eichler (2009) points out that this bivariate approach does not capture Granger’s original definition. Further, it cannot distinguish between causal relationships and correlations between effects of common cause. This can be seen in eq. (A.2) If and have common cause, , and these effects do not always occur simultaneously, then

will provide information about when has occurred and will thus significantly improve the prediction of . A more accurate approach is the multi-variate one, which includes other variables in the model of each time series. Using a vector autoregressive model with variables , now instead of a single variable is a vector representing the measurement of all variables in at time . The system is represented as:

(A.3)

Here is a matrix of coefficients and is a vector of error terms. Using this representation, Granger-causes if at least one of is non-zero.

While this model comes closer to causal inference than the bivariate test does, it has practical problems. Such a model quickly becomes computationally infeasible with even a moderate number of lags and variables. Using a model of order with variables leads to matrices