# Notes on Judea Pearl’s Book Causality, 2nd edition

D. Gueorguiev 9/9/2023

## Preliminary notes on Probability Theory

**Definition** *Conditional Independence*

Let be a finite set of variables. Let be a joint probability function over the variables in , and let , , stand for any three subsets of variables in . The sets and are said to be conditionally independent given if

whenever (1)

In words, learning the value of does not provide additional information about , once we know . (Metaphorically, “screens off” from ).

Eq. (1) is a terse way of saying the following: for any configuration x of the variables in the set and for any configurations and of the variables in and satisfying , we have

(2)

We will use the notation or simply to denote the conditional independence of and given ; thus,

iff (3)

For all values , , such that . Unconditional independence (also called marginal independence) will be denoted with ; that is,

iff whenever (4)

Note that implies the conditional independence of all pairs of variables and , but the converse is not necessarily true.

The following is a (partial) list of properties satisfied by the conditional independence relation .

**Symmetry**:

**Decomposition**:

**Weak union**:

**Contraction**:

**Intersection**:

(Intersection is valid in strictly positive probability distributions)

These properties were named *graphoid axioms* by Pearl and Paz in 1987 and have been shown to govern the concept of informational relevance in a wide variety of interpretations. In graphs, for example, these properties are satisfied if we interpret to mean “*all paths from subset X of nodes to a subset Y of nodes are intercepted by a subset Z of nodes.*”

The intuitive interpretation of the graphoid axioms is as follows:

The *symmetry* axiom states that, in any state of knowledge , if tells us nothing new about , then tells us nothing new about .

The *decomposition* axiom asserts that if two combined items of information are judged irrelevant to , then each separate item is irrelevant as well.

The *weak union* axiom states that learning irrelevant information cannot help the irrelevant information become relevant to .

The *contraction* axiom states that if we judge irrelevant to after learning some irrelevant information , then must have been irrelevant before we learned . Together, the weak union and contraction properties mean that irrelevant information should not alter the relevance status of the propositions in the system;