TECHNICAL REPORT
R-396
August 2014

Graphoids Over Counterfactuals

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
(310) 825-3243 / judea@cs.ucla.edu

Abstract

Augmenting the graphoid axioms with three additional rules enables us to handle independencies among observed as well as counterfactual variables. The augmented set of axioms facilitates the derivation of testable implications and ignorability conditions whenever modeling assumptions are articulated in the language of counterfactuals.

1 Motivation

Consider the causal Markov chain $X \to Y \to Z$ which represents the structural equations:

$$y = f(x, u_1) \tag{1}$$
$$z = g(y, u_2) \tag{2}$$

with u_1 and u_2 being omitted factors such that X, u_1, u_2 are mutually independent.

It is well known that, regardless of the functions f and g, this model implies the conditional independence of X and Z given Y, written

$$X \perp\!\!\!\perp Z \mid Y \tag{3}$$

This can be readily derived from the independence of X, u_1, and u_2, and it also follows from the d-separation criterion, since Y blocks all paths between X to Z.

However, the causal chain can also be encoded in the language of counterfactuals by writing:

$$Y_x(u) = f(x, u_1) \tag{4}$$
$$Z_{xy}(u) = g(y, u_2) = Z_y(u) \tag{5}$$

where u stands for all omitted factors (in our case $u = \{u_1, u_2\}$) and $Y_x(u)$ stands for the value that Y would take in unit u had X been x. Accordingly, the functional and independence assumptions embedded in the chain model translate into the following counterfacutal

statements:

$$Z_{xy} = Z_y \tag{6}$$

$$X \perp\!\!\!\perp Y_x \tag{7}$$

$$Z_{xy} \perp\!\!\!\perp (Y_x, X) \tag{8}$$

Equation (6) represents the missing arrow from X to Z, while (7)–(8) convey the mutual independence of X, u_1, and u_2.[1]

Assume now that we are given the three counterfactual statements (6)–(8) as a specification of some uncharted model; the question arises: Are these statements testable? In other words, is there a statistical test conducted on the observed variables X, Y, and Z that could prove the model wrong? On the one hand, none of the three defining conditions (6)–(8) is testable in isolation, because each invokes a counterfactual entity. On the other hand, the fact that the chain model of Eqs. (1)–(2) yields the conditional independence of Eq. (3) implies that the combination of all three counterfactual statements should yield a testable implication.

This paper concerns the derivation of testable conditions like Eq. (3) from counterfactual sentences like Eqs. (6)–(8). Whereas graphical models have the benefits of inferential tools such as d-separation (Pearl 1988; 2009, p. 335) for deriving their testable implications, counterfactual specifications must resort to the graphoid axioms[2], which, on their own, cannot reduce subscripted expressions like Eqs. (6)-(8) into a subscript-free expression like Eq. (3). To unveil the testable implications of counterfactual specifications, the graphoid axioms must be supplemented with additional inferential machinery.

We will first prove that Eq. (3) indeed follows from Eq. (6)–(8) and then tackle the general question of deriving testable sentences from any given collection of counterfactual statements of the conditional independence variety. To that end, we will augment the graphoid axioms with three auxiliary inference rules, which will enable us to remove subscripts from variables and, if feasible, derive sentences in which all variables are unsubscripted, that is, testable. These auxiliary rules will rely on the composition axiom (Pearl, 2009, p. 229)

$$X_w = x \Longrightarrow Y_{xw} = Y_w \tag{9}$$

which was shown to be sound and complete relative to recursive models (Galles and Pearl, 1998; Halpern, 1998).[3] In the special case of $W = \{\emptyset\}$ the axiom is known as *consistency rule*:

$$X = x \Longrightarrow Y_x = Y \tag{10}$$

and is discussed in Robins (1986) and Pearl (2010).

[1]Rules for translating graphical models to counterfactual notation are given in Pearl (2009, pp. 232–234), based on the structural semantics of counterfactuals. The rules represent the omitted factors affecting any variable, say Y, by the set of counterfactuals $Y_{pa(Y)}$, where $pa(Y)$ stands for the parents of Y in the diagram.

[2]The graphoid axioms are axioms of conditional independence, first formulated by Dawid (1979) and Spohn (1980). Their connections to graph connectivity and to other notions of "information relevance" were established by Pearl and Paz (1987) and are described in detail in (Pearl 1988, pp. 78–133; 2009, p. 11).

[3]The axiom of "composition" was first stated in Holland (1986, p. 968). Its completeness rests on a few technical conditions such as uniqueness and effectiveness (Halpern, 1998).

2 Deriving Testables from Non-testables

In this section we will show that Eq. (3) can be derived from (6)–(8) with the help of (9).

We first note that substituting (6) into (8) yields

$$Z_y \perp\!\!\!\perp (Y_x, X) \tag{11}$$

which is a universally quantified formula, stating that for all z, y, y', x, x' in the respective domains of Z, Y, and X, the following independence condition holds:

$$Z_y = z \perp\!\!\!\perp (Y_x = y', X = x') \tag{12}$$

We next note that, for the special case of $x' = x$, Eq. (12) yields:

$$Z_y = z \perp\!\!\!\perp (Y_x = y', X = x)$$

or, using (10)

$$Z_y = z \perp\!\!\!\perp (Y = y', X = x) \qquad \text{for all } y, z, y', x \tag{13}$$

This can be written succintly as

$$Z_y \perp\!\!\!\perp (Y, X) \tag{14}$$

Our next task is to remove the subscript from Z_y, This is done in two steps. First we apply the graphoid rule of "weak union" (Pearl, 2009, p. 11) to obtain:

$$Z_y \perp\!\!\!\perp (Y, X) \Longrightarrow Z_y \perp\!\!\!\perp X \mid Y \tag{15}$$

Second, we explicate the components of (15) and write

$$Z_y \perp\!\!\!\perp (X, Y) \Longrightarrow Z_y = z \perp\!\!\!\perp X = x \mid Y = y' \tag{16}$$

for all y, z, x, and y'. Again, for the special case of $y' = y$, Eq. (10) permits us to remove the subscript from Z_y and write

$$Z = z \perp\!\!\!\perp X = x \mid Y = y \qquad \text{for all } x, y, z \tag{17}$$

Finally, since the last independency holds for all x, y, and z, we can write it in succinct notation as

$$Z \perp\!\!\!\perp X \mid Y$$

which is subscript-free and coincides with the testable implication of Eq. (3).

To summarize, we have shown that the subscripts in Eq. (11) can be removed in two steps. First

$$Z_y \perp\!\!\!\perp (Y_x, X) \Longrightarrow Z_y \perp\!\!\!\perp (Y, X) \tag{18}$$

and second,

$$Z_y \perp\!\!\!\perp (Y, X) \Longrightarrow Z \perp\!\!\!\perp X|Y \tag{19}$$

Moreover, we see that (3) follows from (8) alone, and does not require the exogeneity assumption expressed in (7).

3 Augmented Graphoid Axioms

In this section we will identify three general rules that, when added to the graphoid axioms, will enable us to derive testable implications without referring back to the consistency axiom of Eq. (10). The three rules are as follows

Rule 1

$$V \perp\!\!\!\perp (X_w, Y_{xw}, S) \mid R \Rightarrow V \perp\!\!\!\perp (X_w, Y_w, S) \mid R \tag{20}$$

Rule2

$$V \perp\!\!\!\perp R \mid (X_w, Y_{xw}, S) \Rightarrow V \perp\!\!\!\perp R|(X_w, Y_w, S) \tag{21}$$

Rule 3

$$V \perp\!\!\!\perp (Y_{xw}, S) \mid (X_w, R) \Rightarrow V \perp\!\!\!\perp (Y_w, S) \mid (X_w, R) \tag{22}$$

Rules 1 and 2 state that a subscript x can be removed from Y_{xw} whenever Y_{xw} stands in conjunction with X_w, be it before or after the conditioning bar. In our example we had $W = \{\emptyset\}$. Rule 3 states that a subscript x can be removed from Y_{xw} whenever X_w appears in the conditioning set. The symbols V, S, R in Eqs. (20)–(22) stand for any set of variables, observable as well as counterfactual.

The proof of these three rules follow the path that led to the derivation of Eq. (18) and (19).

For mnemonic purposes we can summarize these rule using the following shorthand:

Rule 1–2

$$(X_w, Y_{xw}) \Rightarrow (X_w, Y_w) \tag{23}$$

Rule 3

$$(Y_{xw}|X_w) \Rightarrow (Y_w|X_w) \tag{24}$$

4 Deriving Ignorability Relations

Unveiling testable implications is only one application of the augmented graphoid axioms in Section 3. Not less important is the ability of these axioms to justify ignorability relations which a researcher may need for deriving causal effect estimands.[4]

Consider the sentence $Z_x \perp\!\!\!\perp (Y_z, X)$ which may be implied by a certain process, and assume we wish to estimate the causal effect of Z on Y, $P(Y_z = y)$ from non-experimental data. For this estimation to be unbiased, the conditional ignorability $Z \perp\!\!\!\perp Y_z|W$ need to be assumed, where W is some set of observed covariates. Using Axiom (22) we can show that

[4]Reliance on the assumptions of conditional ignorability (Rubin, 1974; Rosenbaum and Rubin, 1983; Holland, 1986), which are cognitively formidable, is one of the major weaknesses of the potential outcome framework (Pearl, 2009, pp. 350–351). Axioms (20)-(22) permit us to derive needed ignorability conditions from other counterfactual statements which are perhaps more transparent.

$W = X$ satisfies the ignorability assumptions and, therefore, adjustment for X will yield a bias-free estimate of the causal effect $P(Y_z = y)$. This can be shown as follows:

$$Z_x \perp\!\!\!\perp (Y_z, X) \implies Z_x \perp\!\!\!\perp Y_z | X$$

(using the graphoid rule of "weak union") and by Rule (22) we obtain

$$Z_x \perp\!\!\!\perp Y_z | X \implies Z \perp\!\!\!\perp Y_z | X$$

We therefore can write

$$\begin{aligned} P(Y_z = y) &= \sum_x P(Y_z = y | X = x) P(X = x) \\ &= \sum_x P(Y_z = y | Z = z, X = x) P(X = x) \\ &= \sum_x P(Y = y | Z = z, X = x) P(X = x). \end{aligned} \tag{25}$$

Equation (25) is none other but the standard adjustment formula for the causal effect of Z on Y, controlling for X.

The process can also be reversed; we start with a needed, yet unsubstantiated ignorability condition, and we ask whether it can be derived from more fundamental conditions which are either explicit in the model or are defensible on scientific grounds. Consider, for example, an unconfounded mediation model in which treatment X is randomized and assume we seek to estimate to effect of the mediator Z on the outcome Y. (The model is depicted in Fig. 1). Operationally, we know that the ignorability condition $Z \perp\!\!\!\perp Y_z | X$ would allow us to obtain

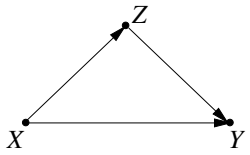

Figure 1: Unconfounded mediation model implying the conditional ignorability $Z \perp\!\!\!\perp Y_z | X$.

the desired effect $P(Y_z = y)$ by adjusting for X, as shown in the derivation of Eq. (25). However, lacking graphs for guidance, it is not clear whether this condition follows from the assumptions embedded in the model; a formal proof is therefore needed. The assumptions explicit in the model take the form

(a) $X_z = X$

(b) $X \perp\!\!\!\perp (Z_x, Y_{zx})$

(c) $Z_x \perp\!\!\!\perp Y_{zx}$.

(a) states that Z does not affect X, (b) represents the assumption that X is randomized, and (c) stands for the no-confounding assumption, that is, all factors affecting Z when X is held

constant are independent of those affecting Y when X and Z are held constant (Pearl, 2009, p. 232, p. 343). These factors stand precisely for the "error terms" that enter the structural equations for Z and Y, respectively, hence they have clear process-based interpretations, and avail themselves to plausibility judgments.

To show that the desired ignorability condition $Z \perp\!\!\!\perp Y_z|X$ follows from (a), (b) and (c), we can use Rule 3 (Eq. 22) as follows. First, the standard graphoid axioms dictate

$$X \perp\!\!\!\perp (Z_x, Y_{zx}) \;\&\; Z_x \perp\!\!\!\perp Y_{zx} \Rightarrow Z_x \perp\!\!\!\perp Y_{zx}|X$$

Next, applying Rule 3 twice, together with $X = X_z$, gives

$$Z_x \perp\!\!\!\perp Y_{zx}|X \Rightarrow Z \perp\!\!\!\perp Y_{zx}|X \Rightarrow Z \perp\!\!\!\perp Y_{zx}|X_z \Rightarrow Z \perp\!\!\!\perp Y_z|X_z \Rightarrow Z \perp\!\!\!\perp Y_z|X$$

which yields the desired ignorability condition.

These derivations can be skipped, of course, when we have a graphical model for guidance. The adjustment formula (25) could then be written by inspection, since X satisfies the back-door condition relative to $Z \to Y$. However, researchers who mistrust graphs and insist on doing the entire analysis by algebraic methods, would need to use Rules 1–3 to justify the ignorability condition from assumptions (a), (b), and (c).

5 Conclusions

Rules 1-3, when added to the graphoid axioms, allow us to process conditional-independence sentences involving counterfactuals and derive both their testable implications, as well as implications that are deemed necessary for identifying causal effects. We conjecture that Rules 1–3 are *complete* in the sense that all implications derivable from the graphoid axioms together with the consistency rule (18) are also derivable using the graphoid axioms together with Rules 1–3.

Augmented graphoids are by no means a substitute for causal diagrams, since the complexity of finding a derivation using graphoid axioms may be exponentially hard (Geiger, 1990). Diagrams, on the other hand, offer simple graphical criteria (e.g., d-separation or back-door) for deriving testable implications and effect estimands. In reasonably sized problems, these criteria can be verified by inspection, while, in large problems, they can be computed in polynomial time (Tian et al., 1998; Shpitser and Pearl, 2008). The secret of diagrams is that they embedd all the graphoid axioms in their structure and, in effect, pre-compute all their ramifications and display them in graphical patterns.

Acknowledgment

Sander Greenland and Jin Tian provided helpful comments on an early version of this note. This research was supported in parts by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211 and #IIS-1018922, and ONR #N000-14-09-1-0665 and #N00014-10-1-0933.

References

Dawid, A. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* **41** 1–31.

Galles, D. and Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundation of Science* **3** 151–182.

Geiger, D. (1990). Graphoids: A qualitative framework for probabilistic inference. Ph.D. thesis, University of California, Los Angeles, Department of Computer Science.

Halpern, J. (1998). Axiomatizing causal reasoning. In *Uncertainty in Artificial Intelligence* (G. Cooper and S. Moral, eds.). Morgan Kaufmann, San Francisco, CA, 202–210. Also, *Journal of Artificial Intelligence Research* 12:3, 17–37, 2000.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81** 945–960.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2010). On the consistency rule in causal inference: An axiom, definition, assumption, or a theorem? *Epidemiology* **21** 872–875.

Pearl, J. and Paz, A. (1987). GRAPHOIDS: A graph-based logic for reasoning about relevance relations. In *Advances in Artificial Intelligence-II* (B. D. Boulay, D. Hogg and L. Steels, eds.). North-Holland Publishing Co., 357–363.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling* **7** 1393–1512.

Rosenbaum, P. and Rubin, D. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika* **70** 41–55.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688–701.

Shpitser, I. and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* **9** 1941–1979.

Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic* **9** 73–99.

Tian, J., Paz, A. and Pearl, J. (1998). Finding minimal separating sets. Tech. Rep. R-254, University of California, Los Angeles, CA.