International Statistical Review (2002), **70**, 2, 191–214, Printed in The Netherlands
© International Statistical Institute

Seeing and Doing: the Concept of Causation

Dennis V. Lindley

"Woodstock", Quay Lane, Minehead, Somerset, TA24 5QU, UK. E-mail: thombayes@aol.com

Summary

This note is an extended review of the book by Judea Pearl (2000) on causality, in which the basic concepts therein are explained in a form that statisticians will hopefully appreciate, including some comments on their relevance to inference and decision-making.

Key words: Causality; Multivariate distribution; Causal mechanism; Association; Regression; Simpson's paradox; Directed acyclic graph; Structural equation; Covariate; Counterfactual.

1 Introduction

Statisticians rarely refer to causality in their writings and, when they do, it is usually to warn of dangers in the concept. Thus Speed (1990) says "Considerations of causality should be treated as they have always been treated in statistics: preferably not at all but, if necessary, then with very great care". Cox & Wermuth (1996) remark "We did not in this book use the words *causal* or *causality* .... Our reason for caution is that it is rare that firm conclusions about causality can be drawn from one study". Lindley & Novick (1981), in a paper discussed in chapter 6 of this book, say that "causality, although widely used, does not seem to be well-defined" and therefore reject the term. This avoidance by statisticians is strange because the term is widely used both by scientists and laymen, who presumably base their use of it on inference from data, which topic is the preserve of statisticians. Their preferred form of inference is association, captured in the techniques of correlation and regression, which is a weaker idea than causation, and therefore leads to weaker results. Causality also has a stability that association does not; a stability that is reflected in a basic assumption made in this book.

In recent years, there have been several studies, outside the statistical profession, about the notion of cause and, whilst progress has been made in understanding the topic, there has been no totally-satisfactory account. What has been especially lacking is a calculus of causality so that users can operate with the concept just as confidently as they can with probability through its own calculus. In this book, Pearl presents a description of causality, with its accompanying calculus, through the concept of a causal mechanism, within which it is possible to perform manipulations that lead to interesting and effective, new results. This is a major contribution to our appreciation of causality, not so much at the philosophical level, but more as an important tool for calculation. Statisticians should study the work carefully with the expectation that it will alter their attitude to causation and enable them to incorporate causal concepts into their work. The present note tries to summarize the ideas in a form that might be appreciated by statisticians, and will hopefully encourage them to read the book themselves and experiment with the causal calculus. The note concludes with some comments on the form of the book and its contents.

2 Multivariate Distributions

Whatever is meant by causation, it is clear that it refers to a relationship between quantities, a change in one being the cause of a change in another. Ordinarily there is some uncertainty involved, as when it is said that the application of fertilizer will cause the crop yield to increase, the actual yield that might result being uncertain. The tool for the study of uncertainty is probability, with its familiar calculus, so it is there that the study begins. Pearl uses the Bayesian interpretation of probability, as a degree of belief, but this should not deter statisticians who prefer to think in terms of frequency, because most of the book is concerned with manipulations where the concept of a population can equally be used. Indeed, in examples, he mostly slips from frequency to belief with scarcely a murmur. The study of causality therefore begins with uncertain quantities, often called random variables, and their probability specification, which can be expressed in several forms. One form is a joint distribution but a preferred method is to place the quantities in some order and use a sequence of univariate distributions. Thus, in a simple example with three quantities, u, x, y, when placed in that order, may have their joint distribution written as

$$p(u)p(x|u)p(y|u,x) \tag{1}$$

where the distribution, here expressed through a density, of each quantity is conditional on the values of all the quantities that have preceded it in the chosen ordering. The first stage in the construction of a causal mechanism is to select an ordering of all the quantities and to write down their probability structure in terms of that ordering.

Whilst the introduction of an order into a set of quantities is often a useful tool in the construction of a joint distribution, any order can be used, and it is easily possible to switch from one to another. Thus, if the order above, u, x, y, is replaced by x, u, y, the joint distribution becomes

$$p(x)p(u|x)p(y|x,u), \tag{2}$$

which is equivalent to (1), passage between them being accomplished by means of the probability calculus, here using the product rule on the first two densities in both (1) and (2). This is fine when discussing association, for if x is associated with u, then equally u is associated with x; but it is unsatisfactory for causation, for if x is a cause of u, it is emphatically not true that u is a cause of x. Also $p(u|x)$ is easier to think about than $p(x|u)$. Consequently a causal mechanism includes not just a joint probability distribution of a set of quantities but also a specific ordering of them. To anticipate, if x is a cause of u, then x precedes u in the ordering and (2) might be relevant, whereas (1) would not be. It is important to notice that the procedures to be described later do not remain invariant under a change of order but are heavily dependent on the order selected.

Statisticians have recognized the relevance of order in some cases; for example in passing from correlation between y and x, to the regression of y on x, where the former is unaltered when the quantities are interchanged but the latter is not. There is a further point about regression in that there is a distinction between $p(y|x)$ as part of a joint distribution, where it appears together with $p(x)$, and $p(y|x)$ where x loses its uncertain status and has been selected. Pearl expresses the distinction as one between seeing X to be equal to x, and controlling X to be x, where upper-case has been used to denote the quantity (variable) and lower-case to describe its value. We will follow him and use the rather imprecise, but most useful, notation $p(y|x)$ in the former case with x random and $p(y|do(x))$ in the latter with x controlled. When it is useful to emphasize the distinction, he writes $p(y|see(x))$ instead of $p(y|x)$. Much of the book is devoted to the 'do' operator and the mechanisms for handling it, so providing the calculus of causality mentioned earlier.

It has long been recognized that there can exist real differences between $p(y|see(x))$, with x uncertain, and $p(y|do(x))$ with x selected. A famous illustration is with x size of foot and y size of hand. The binding of feet, practised in some societies to lessen x, has no effect on size of hand, so that $p(y|do(x))$ could be $p(y)$, certainly not $p(y|x)$. To anticipate, in this example, x and y

have a common, genetic cause, u, so that change in x does not influence u and the link between x and y is broken. $p(x|u)$ has a stability that $p(u|x)$ does not possess. Equally there are cases where $p(y|x) = p(y|do(x))$, as when $X = 1(0)$ if a treatment is (is not) adopted in some randomized, clinical trial; the behaviour in the trial being extended to advise about treatment, believing that if the treatment is taken, the results will correspond to those in the trial. One might say that the treatment caused the patients to recover. The distinction also arises in decision analysis in comparing decision and random nodes. There can be a real difference between a decision concerning X and a random occurrence of X; the former calling for maximization of utility, the latter for calculation of expected utility.

3 Causal Mechanisms

With these preliminaries, it is now possible to see what is meant by a causal mechanism. We illustrate with the case of three quantities. As explained, they are first placed in some order, say u, x, y, and that order used to describe their joint distribution as

$$p(u)p(x|u)p(y|u,x). \tag{1}$$

What happens if x is controlled, or selected, and replaced by $do(x)$? Clearly x now has no uncertainty, and $p(x|u)$ is no longer relevant, but it is not clear what happens to the remaining densities in (1). Pearl makes two assumptions:

(a) $p(u)$ is unaffected by the control of x, and

(b) in $p(y|u,x)$, x is replaced by the value selected by the control but otherwise remains unaltered. In the 'do' notation, $p(y|u, do(x)) = p(y|u,x)$.

As a consequence of these two assumptions, (1) is replaced by

$$p(u)p(y|u,x), \tag{3}$$

where it is understood that the value of x is that selected by the control. Pearl refers to the individual, univariate components in the product as autonomous processes being unaffected by control of one quantity, except insofar as everywhere else that quantity is set equal to its controlled value.

In general, assumption (a) says that if a quantity is controlled, then the probabilities of all quantities that precede it are unchanged. (b) says that the only change to all quantities that succeed it is merely to replace the general x by the controlled value, so that, in the example, the regression of y on (u,x) is the same whether x is seen or is controlled. Although this regression is supposed unaltered, that of y on x alone is changed, for when x is random, (1) applies and

$$p(y|x) = \int p(u)p(x|u)p(y|u,x)du \,/ \int p(u)p(x|u)du\ ,$$

whereas if x is selected, (3) is operative and we reach the different conclusion that

$$p(y|x) = \int p(u)p(y|u,x)du. \tag{4}$$

Similarly it can easily be seen that the ordering of the quantities is material, for if, in place of u, x, y, the order x, u, y is employed, (2) is relevant and a third regression of y on selected x is obtained:

$$p(y|x) = \int p(u|x)p(y|x,u)du, \tag{5}$$

which is the familiar rule of probability in which the conversation is extended from (x, y) to include u, unlike (4).

An impressive illustration of these ideas is provided by the case where x is a treatment, y a response and u a covariate which may be associated with both x and y. Clearly x influences y, and so precedes it in the ordering, but whether u influences x, or the reverse, vitally affects the efficacy of the treatment on the response, judged by $p(y|x)$. To see this suppose x influences u and the order x, u, y is relevant; then from data on the three quantities, the regression $p(y|x)$ given by (5) applies. Hence if $x = 1(0)$ when the treatment is (is not) applied, and $y = 1(0)$ if the outcome is beneficial (harmful), the treatment may be judged beneficial if

$$p(y = 1|x = 1) > p(y = 1|x = 0). \tag{6}$$

In the other case, where u influences x and the order is u, x, y, the efficacy of the treatment is again determined by $p(y|x)$, which now involves $p(y|u, x)$ and $p(u)$, from (4). In particular, the treatment may be judged harmful when

$$p(y = 1|u, x = 1) < p(y = 1|u, x = 0) \tag{7}$$

for all u. It is well-known that (6) and (7) can happen simultaneously; the phenomenon is known as Simpson's paradox, though, as Pearl points out, it was known to Karl Pearson. Consequently the paradox is elegantly resolved by noticing whether x influences u or vice versa, which is reflected in the mathematics by the order selected.

4 Alternative Approaches

The discussion has here been presented in terms of probability but there are two alternative tools that can usefully be used. The first is to employ a directed, acyclic graph (DAG) in which each quantity is represented by a node and a link between two nodes has a direction from x to y iff x precedes y in the ordering. DAG's are conceptually very useful but their advantages extend beyond that. If $(x_1, x_2, \ldots, x_n)$ is a set of uncertain quantities in that order, the general term in the joint probability specification will be $p(x_j|x_1, x_2, \ldots, x_{j-1})$ in which x_j depends on its predecessors. It frequently happens that not all of the predecessors enter into the condition, so that only some appear, called the parents of x_j, written px_j. Equivalently, given px_j, x_j is independent of its other predecessors, when the links in the DAG between x_j and its non-parents may be omitted. The methods using the 'do' operator previously described, in particular the two basic assumptions (a) and (b), do not involve independence; nevertheless, independence conditions do enormously simplify the calculations, so that Pearl develops methods of handling DAG's that occupy a significant portion of the book and are valuable in handling complex situations.

An alternative description is offered in terms of functions, in which one writes $x_j = f_j(px_j, \varepsilon_j)$, where f_j is a deterministic function of the parents of x_j and of a random disturbance ε_j, so that the probability structure passes from the x's to the ε's. Econometricians are fond of this sort of representation, under the term 'structural equation modelling', and Pearl has valuable comments on the practice. Personally I find the approach confusing because if, reverting to the case of three quantities, $y = f(x, \varepsilon)$ with x preceding y, it does not make sense, within the 'do' calculus to invert and write x as a function of y, corresponding to y preceding x. However, examples with $y = f(x)$, and simultaneously $x = g(y)$, do appear (1.42 and 1.43) with resulting confusion. It is not always clear what functional operations are permissible, whereas it is always clear what probability reversals are available; though this preference may only reflect my own experiences.

5 Compliance, Counterfactuals

In constructing a model with a sensible ordering that is judged to satisfy the assumptions (a) and (b) above, it is often necessary to introduce quantities that are unobserved. In the case above with x treatment and y response, the covariate u may not be observed, yet may influence y and perhaps x. The influence on x is often removed by choosing x at random, implying x is independent of virtually everything. Pearl therefore distinguishes between quantities that are observed and those that are not;

where a major task is to determine whether, having used the 'do' operation to obtain the required result, exemplified by (4), the probabilities needed can be obtained from observations. There is a powerful example of his techniques in §8.2, dealing with a problem that is firmly within the ambit of statistical inference. Here x, as before, is an assigned treatment, y a response and u a covariate; but there is an additional quantity z which refers to the treatment actually used, so that comparison between x and z indicates whether the subject, to whom the probabilities apply, complied with the treatment x assigned. It is supposed that u influences z and y, but not x, assigned at random, and the order is x, u, z, y. It is not difficult to evaluate $p(y|do(z))$ but can it be determined from observations on x, y, z, but not u? It cannot but in a masterly series of calculations within the probability calculus, it is shown that useful bounds to the probability can be evaluated.

Pearl makes an important contribution to the study of counterfactuals. In the case of binary quantities, x and y, taking values 'true' and 'false', a counterfactual is a probability statement about the truth of y, had x been true, when it is known that y had been false when x was false. An example arises in medical litigation when the complaint is made that A would be alive if the doctor had operated, when he had not and A died. The connection with causality is clear; the failure to operate may have caused A's death. The claim is made (p. 218) that scientists are concerned with counterfactuals rather than predictions. To illustrate the method of solution, consider, as before, three quantities u, x, y in that order, where $p(y|x)$ is given by (4). Now suppose additional evidence e becomes available that 'x is false and y did not occur'. The effect on the probability structure is easily handled by Bayes's theorem to revise $p(u)$ to $p(u|e)$. Consequently if we set x to be true, the regression (4) becomes

$$p(y|x) = \int p(u|e)p(y|u, x)du$$

and a statement about y can be made were x true. This method depends on the introduction of the covariate u and that it satisfies, in conjunction with x, y, the assumptions (a) and (b). An extension of these ideas leads to probability statements about causes.

6 Commentary

The above is an inadequate summary of the book, omitting the important calculations and applications but, I think, including the key ideas. I now turn to some comments on the material. All the procedures developed depend on the ability to order a set of quantities, observed and unobserved, in such a way that assumptions (a) and (b) obtain; that is, setting one quantity to a chosen value has no effect on its predecessors and only effects its successors by replacing uncertain x by chosen x, leaving the autonomous, conditional distributions unaltered, $p(y|u, x) = p(y|u, do(x))$ in our illustrative example. Such a system is termed a causal mechanism. The whole edifice constructed in the book depends on the validity of (a) and (b), and it is interesting that Pearl does not refer to them as assumptions, but as definitions (7.1.3 and 7.1.4). The change is not, I feel, purely linguistic. Whenever a model is used, the presumption is being made that it is relevant to the real world and it behoves one to make such checks as are possible to assess its relevance, perhaps even its truth. Consequently it would appear essential in any application to check that the two assumptions are reasonable, for there are cases where they are not: seeing someone dead is radically different from making someone dead. For a more relevant example, a chemical engineering study may have produced data that demonstrate that the temperature of the process in the vessel affects the quality of the final product, with the consequent conclusion that the process should include a temperature control. However, this inclusion may involve considerable changes to the vessel and it would be unreasonable to suppose $p(y|x)$ from the study is the same as $p(y|do(x))$, even when account is taken of the covariates, as (b) demands. One response to this objection might be to say that the model used in the study is inadequate and that a fuller one that incorporates possible changes to the engineering is required, but this presents further difficulties which can be illustrated on the compliance model mentioned earlier. Recall that here u was an unobserved covariate that influenced both compliance z and response y, but not x, with order x, u, z, y where x is the treatment assignment. With the assumption that, given (u, z), y is independent of x, the causal mechanism requires that $p(y|u, z)$ is unaffected by changing to $do(z)$, a requirement that may be impossible to verify if u is not merely unobserved but unstated. A similar difficulty with an unspecified covariate arises in the familiar example connecting x, smoking, with y, lung cancer, where z, tar deposit in the lungs, acts as an intermediary and u is an unspecified, genetic covariate; the order being u, x, z, y, Figure 3.5. It is there supposed that $p(z|u, x) = p(z|x)$, a result that it is hard to see being verified.

My first reaction to the strong assumptions (a) and (b), and hence to the whole edifice here constructed, was one of extreme scepticism, but increased exposure to the ideas and to real problems has lessened this appreciably to one where my main reserve concerns difficulties in checking the reasonableness of the assumptions. Extensive use of causal ideas by scientists and laymen suggests that we do have an appreciation of order, that x influences y, rather than y influences x, and that we do feel that some structures are autonomous enough to resist interference by control. Consequently causal mechanisms may be easier to come by than one might initially think. What is also abundantly clear is that if the assumptions are made, then powerful results follow. Also, to be fair, Pearl, in chapter 2, does have useful things to say about the construction of causal mechanisms. It appears to be reasonable to accept the assumptions and explore the rich consequences that flow from them, rather than engage in too much speculation about their validity, which can be considered if the consequences appear unsound. There will always exist multivariate situations which are not amenable to this treatment, often because no causal structure can be seen and the quantities have, as a result, no natural order or DAG.

The calculus is especially apposite for a Bayesian. In the personalistic approach adopted by an individual 'you', you are free to assess $p(y|see(x))$ and $p(y|do(x))$ in any way, as separate evaluations, without violating coherence. As someone has said, the only way to know what might happen were you to control x, is to control x; merely seeing x can be insufficient. The causal mechanism permits you coherently to identify the two probabilities under some circumstances, but not others. For example, in (1), $p(y|u, x)$ is, for you, the same for uncertain x or selected x; but as we saw, this is not true for $p(y|x)$, (4) obtaining when x is selected, different from the usual form when x is uncertain. As a result the 'do' calculus fits neatly with its probability companion.

Statisticians will, I think, find the book difficult reading; I certainly did. One reason is that Pearl pays little attention to the relationship between data and the models discussed; an exception is §8.5. As a result it may be hard for an applied statistician, used to data analysis, to appreciate the arguments presented. Nevertheless, the principal aim of the book is to develop some powerful machinery and that to deal, in addition, with the associated inference problem would both distract from the main task and make an already long book even longer. There is an epilogue, in the form of a public lecture, which is easy reading but is understandably too superficial to give a sound understanding of the relevance of the book to statisticians.

Statisticians of a more mathematical bent will have a different difficulty for, instead of developing his ideas in a logical order, the author scatters his pearls without the necklace's thread. It is only in chapter 7 that the anticipated development appears, by which time the reader is expected to have absorbed enough ideas to dispense with the illustrative examples that even the most mathematical treatment demands. There is a further problem that the definitions are not always well-expressed; a feature which is not unknown amongst engineers who have high manipulative skills in mathematics but are not so good at stating the basic concepts. An example is provided on p.189: "the condition of unbiasedness (Definition 6.2.1) does not imply the modified criterion of Definition 6.3.2". This appears to be the first usage of the term 'unbiasedness', which does not occur in the index (which is poorly done, even the term 'cause' does not appear!) and is not present in either of the definitions mentioned. This attitude towards definitions is in contrast to his dismissal (p. 178) of de Finetti's precise definition of 'exchangeable' as meta-statistical, a criterion that could be applied to some in the book. Another criticism is the plethora of terms, many of which may be unnecessary, like probabilistic, statistical and causal parameters (§1.5) which are rarely used.

Some difficulties in presentation may arise because of the modern haste to publish. Darwin only wrote about evolution after he had thought about the ideas for decades and, as a result, we have a masterpiece of both science and literature. Today social pressures almost force early publication before the ideas have had time to settle, the result being books that are untidy and often not properly understood, even by their authors. Pearl might also learn from Darwin that a clear statement of one's position is all that is needed, a brilliant idea can speak for itself, and attacks on others are not needed. (Did you know that my generation of statisticians is 'tormented'?) Another difficulty is that the author repeats himself, sometimes using different terms. For example, the condition $p(y|x) = p(y|do(x))$ is referred to as unbiasedness (6.10) or as exogenous (7.46); page 206 echoes page 37. The repetition may be the result of the book not being written afresh but formed by putting papers together.

The comparison with Darwin may be exaggerated but it is surely true that this is an important book that ought to be read by statisticians, if only to appreciate when seeing and doing are comparable, and to explore the wide range of consequences that follow from supposing that they do.

References

Cox, D.R. & Wermuth, N. (1996). *Multivariate Dependencies—Models, Analysis and Interpretation*. London: Chapman and Hall.

Lindley, D.V. & Novick, M.R. (1981). The role of exchangeability in inference. *Ann. Statist.*, 9, 45–58.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Speed, T.P. (1990). Complexity, Calibration and Causality in Influence Diagrams. In *Influence Diagrams, Belief Nets and Decision Analysis*, pp. 49–63 Eds. R.M. Oliver and J.Q. Smith. New York: Wiley.

Résumé

Cet article part d'une critique du livre de Judea Pearl (2000) sur la causalité, dans lequel les concepts de base sont expliqués sous une forme que les statisticiens devraient apprécier. Il comprend aussi des commentaires sur le rapport de ces concepts avec l'inférence et la prise de décisions.

Mots clés: Causalité; Distribution multivariée; Mécanisme causal; Association; Régression; Paradoxe de Simpson; Graphique acyclique; Equation structurelle; Variable explicative.