Causal Analysis and Statistics: A Social Sciences Perspective

Elja Arjas

Human beings exhibit an almost obsessive urge to mold empirical phenomena conceptually into cause–effect relationships. This tendency is, in fact, so strong that it sometimes comes at the expense of precision and often requires the invention of hypothetical unobservable entities (such as ego, elementary particles, and supreme beings) to mold such theories into causal schemata (Judea Pearl).

Causal Explanation as a Scientific Ambition

Causal reasoning is indeed a basic element of our thinking, both in setting up scientific theories and as part of our everyday reasoning. This is true in spite of the fact that, as has been clear since the work of Hume more than 250 years ago, there is ultimately no way of knowing whether causality, in an objective sense, actually exists. In a deterministic world-view causality is not needed. Indeed, there is no room for it: everything evolves along a predetermined trajectory, without needing a real reason, as there are no alternatives.

All causal concepts are built on the idea of comparing what really happened, under certain circumstances, with what might have happened under some particular different circumstances. Thus the causal effect of some action on an individual – supposing that the action actually took place – would be defined as the difference between the true outcome and one that would have followed from a counterfactually different action. But in such a comparison the latter 'might have, but did not' alternative is a mental construct which will exist only in our minds. As a consequence, taking this literally, individual causal effects can never be measured in the physical world in which we live.

Even if one were willing to accept the idea of individual-level causal effects, there are further conceptual problems to tackle. How is one supposed to know what 'would have happened' to an individual? This question seems particularly relevant in empirical research arising in behavioural and social sciences, which aim to study the behaviour of conscious individuals or of groups consisting of such individuals. Should the construction of the individual causal effect be based on what a real person belonging to the study population himself/herself knows to be true, or to have been true, or only on what a sociologist or statistician can establish from the data? The former type of knowledge can perhaps be approached in qualitative analysis based on individual interviews. But an ordinary statistical analysis of quantitative and attribute data will easily lead to generic conclusions and interpretations about cause and effect that have little contact with the true behaviour and reasoning of the individuals considered.

These difficulties create a genuine paradigmatic dilemma: on one hand, causality is thought to be an essential element of all truly scientific reasoning, yet all attempts to formalize it are based on mental constructions which, by definition, do not have objectively defined counterparts in the world we live in. This results in a tension between what a scientist would ideally like to do, in the way of 'molding empirical phenomena conceptually into

cause–effect relationships', and what can actually empirically be accomplished, even in principle.

From Association to Causation

Design Issues

For the most part, statisticians have been conspicuously silent about what their discipline could do to relieve such tension. Statistical textbooks typically only issue the warning that correlation and association are not the same as causality, but do so without being more specific about what the difference might be. The role of statistics is then restricted to demonstrating, by means of hypothesis testing, that some parameters of interest have values which are different from zero in the sense of statistical significance. As most scientists do not recognize that merely reporting such statistically significant associations found in the data would be a relevant goal for their study, but would prefer to use the language of causality, they have to use some other means to bridge the gap between where the statistical analysis left them and the achievement of a causal conclusion. Sound arguments for doing this are then essentially contextual, based on the analyst's knowledge about the subject-matter. This is by itself a valid approach (cf. Freedman, 1999), and it is only fair to say that no formal statistical machinery can compensate for sensible argumentation based on the subject-matter. This view can be contrasted with some recent ideas coming from the artificial intelligence community, which suggest that it would be possible to decide whether some observed dependence is causal by applying a formal probabilistic framework.

Nevertheless, it makes sense to look for statistical formulations and methods that would correspond more closely to the notion of causality, as it is commonly perceived, than merely testing for statistically significant associations. Expressed in simple terms, the basic idea about causal dependence is that by choosing an action one will also influence the outcome. In designed experiments actions can really be chosen, and could therefore in principle have been chosen differently. But in the social sciences designed experiments are a rare exception, and observational studies are the rule. Rather than the investigator being a free agent who, in a prospective manner, can choose between alternative treatments to be given to the subjects under study, data are typically collected retrospectively, without any ability to control the lives of the participating individuals through planned interventions. Instead, the process of data collection can often be controlled, and therefore the issue is whether there exists a sampling design which in principle would help us to settle a causal claim.

Indeed, as a substitute for a proper experiment, one can sometimes entertain the idea of a pseudo-experiment, which from a causal analysis perspective has the same important characteristics as an actual designed experiment. As an alternative to actual control, one can then speak about the choice of circumstances in which the sampling is made (Rosenbaum, 1999). For example, if a company is closed down during a recession and its entire workforce is made redundant, one can study the effect of unemployment on its former employees without thinking that those who lost their jobs did so because they appeared to be the least fit for the occupations in question (Martikainen and Valkonen, 1998). Doubts about the presence of such selection mechanisms are an example of confounding, which, if present, has the potential to contaminate and hence ruin the whole logic of a causal analysis. Another interesting example of a pseudo-experiment design is the study of the effects of the loss of a spouse or a child in a car crash (Lehman *et al.*, 1987, Rosenbaum, 1999).

Causality and Time

It can be said to be axiomatic to any notion of causality that it can act only forwards, that is, a cause must precede its effect in time. Given this, it is quite striking that most probabilistic/statistical formulations of causal dependence suppress time completely. But this simple requirement of time ordering is not the only aspect of causality where time is important. For example, a therapy may itself be risky, and therefore its effect on a patient may well be negative if only a short follow-up time is considered. But when the follow-up time becomes longer, the direction of the effect may be reversed as the benefits of the therapy become clear. Further issues relating to time are that it may matter when the therapy was given, and also what had happened in the past before it was given.

These aspects relating to time can be conveniently dealt with by employing the general modelling framework of event-history analysis (see e.g. Blossfeld *et al.*, 1989; Arjas and Eerola, 1993; Eerola, 1994; Parner and Arjas, 1999). One will then consider an ordered sequence $0 < T_1 < T_2 < T_n < \ldots$ of event times and a corresponding sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$, where X_n is a description of the event that took place at time T_n. Furthermore, for considering causal questions, it is convenient to split such descriptions X_n into two parts. Writing then $X_n = (Z_n, A_n)$, Z_n typically denotes time-dependent covariates of the subjects considered in a follow-up at time T_n and A_n is an action taken at that time. The reason for this separation in the event description is that while Z_n is thought to represent measured characteristics, A_n signifies the result of a decision which, at least hypothetically, is optional and can be chosen according to 'free will'. The variables A_n are therefore the link to counterfactual reasoning: keeping the past history fixed at its observed value, we are thinking of how different actions A_n 'now' will influence predictions concerning the value of a future response, say Y. In principle, Y can be any random variable included in the model, such as some function of the realized event history.

The Issue of a Confounded Analysis

A fundamental requirement in all empirical studies directed towards causal conclusions is that potential confounding variables must be controlled. The intuition behind this is that an action will have a causal effect on a response if the response changes when the action changes 'under circumstances in which everything else that is relevant to the outcome remains the same'. In other words, fixing the values of potential confounders is an attempt to isolate the causal mechanism from possible alternative ways of explaining the observed response.

Although the need to control for potential confounders is generally accepted, it is surprising how little has been done towards developing a statistical framework within which a corresponding hypothesis could be formulated in an exact manner. Important contributions in this direction have been made particularly by Rubin (1974) and Robins (1997). Rubin introduced the concept of 'ignorable treatment (assignment)' to cover situations in which there is a single action. A key idea in his definition was his postulation of the existence of a collection of potential outcomes Y_a, one for every possible action a. The ignorable treatment hypothesis states then that the choice of the action has to be independent of the values of the potential outcomes Y_a. Robins extended this postulate to sequential designs, then calling it 'no unobserved confounders'. Although both postulates seem to be relatively straightforward conditional independence statements, it is nevertheless quite hard to understand intuitively what independence from a collection of potential outcomes would really mean. It appears to require that the values of such potential outcomes would already have been realized before the action was taken, which seems to run counter to the axiomatic time ordering between cause and effect. Moreover, considering the fact that at most one of such potential outcomes can ultimately be realized, it seems wasteful to introduce an entire collection of random variables, or their joint distribution, when in fact only one will ever be needed. This is akin to the confusing tradition of introducing latent lifetimes in competing risks theory. Finally, it goes against the common practice in probability theory of introducing 'conditional random variables' when what is needed can be formulated in terms of conditional probabilities.

An alternative formulation to the ignorable treatment hypothesis was recently given by Parner and Arjas (1999). The general formulation of that hypothesis is technically somewhat involved, but the intuition behind it is quite straightforward and closely resembles that underlying the definitions of Rubin and Robins. Stated simply, we need a requirement that the choice of an action does not depend on unobserved factors that themselves could influence the considered outcome. But making such a condition precise is actually quite subtle. Clearly, everything in the past that is fixed by the observed data remains fixed at the observed value, no matter what action might be chosen next. Therefore, the use of a statistical model to control explicitly for more variables and account for their influence on the outcome, provides a systematic way of decreasing the number of such potential confounding variables.

The next issue is what can be said about variables whose values were not observed. Stating a condition that for 'everything that has not been observed' seems far too imprecise to provide a basis for systematic causal analysis and so it is necessary to restrict such considerations to the relevant causal field of variables (Mackie, 1965). This concept, which is analogous to the 'small world' concept of Savage (1972), once again emphasizes the importance of the analyst's subjective choices about what variables would be meaningful in the considered context. Here knowledge of the subject-matter becomes crucial, because one can never be sure that all the factors that might contribute to the outcome have been included in the causal field under study. The larger the field, however, the more problematic it becomes to assume that the property stated above of conditional independence between each action and the unobserved variables in the causal field would hold. (Note that empirical verification from the data is out of the question here because these variables were not observed.) Therefore, rather than asking for convincing objective criteria for a causal claim, the analyst can at best ask that the arguments he or she presents in support of it are accepted and shared by others.

Interpretation of Probability

Randomness and probability come into causal analysis, in essence, because as analysts of the data we are generally unable to make accurate predictions of the outcome variable based on the information that is available. (This does not by itself imply a firm position about whether the phenomenon being studied should be viewed as being deterministic or stochastic. In the latter case, accurate prediction would of course be impossible even in principle.) Perhaps somewhat unfortunately, statisticians, and even philosophers, have sharply divided opinions about what probability means. Such differences in interpretation are then also reflected in the statistical paradigms which are followed in analysing data, as well as in the way the empirical evidence resulting from such analysis is related to a causal claim.

The classical probability concept based on symmetry is obviously too limited to be of much practical use in serious scientific study involving causal argumentation. But then, what are the alternatives? Blossfeld and Rohwer (1995) suggest that probability be viewed primarily as a propensity of the object or individual being considered. This notion of probability has the obvious merit that it can in some sense be thought to have an objective existence, so that making causal statements which are based on comparing such probabilities gives the impression that they share the same characteristics and ideals as similar statements made in the physical sciences. But in social sciences, does it really make sense to take such an outsider's view of the behaviour of our fellow humans? I am somewhat doubtful that it does. In particular, the propensity notion seems to ignore the fact that in social sciences one is studying individuals who, in ways that are not made explicit in the data, are conscious about their own behaviour and, as is commonly thought, may choose their actions according to their own free will.

A view shared by most practising statisticians is that probability should be understood as a relative frequency in an idealized infinite population. It is then legitimate to speak about an average effect of some action, such as the introduction of a new tax law, on the population, this being equal to the contrast between the expected values of the response 'before' and 'after'. But if such an interpretation is followed, one is not really saying anything about the effects of the action on the individual members of the population, only on the collective. The frequency interpretation of probability might be an adequate measure in describing how a fertilizer influences the yield of a certain crop, but its focus on populations rather than on their individual members makes it in my view less appropriate for describing causal dependence in a behavioural or social sciences context.

A very different view is that probabilities do not have objectively defined existence at all; or that even if they do, as might be the case in quantum mechanics, such an interpretation would not be relevant to the study of cause and effect in a social science context. Stated simply, randomness means 'uncertainty', and probability is ('My') subjective quantification of the degree of such uncertainty. More accurate knowledge about individual circumstances will generally decrease the degree of randomness, but ultimately accounting for some

amount of randomness will be necessary in nearly all modelling of empirical phenomena. This seems particularly true in the social sciences, where the information basis can never be complete enough so as to make the prediction of individual events accurate.

Adopting an openly subjectivist view of probability obviously offers the researcher a great deal of freedom, since numerical values given to probabilities are no longer viewed as objective claims about Nature, including humans, but as statements of what 'I', as an investigator, think. But it would be unfair to say that such a subjectivist position implies that one is making claims that are arbitrary. Actually, the opposite is true in the sense that within such a framework the methods of statistical inference, being completely based on a probability calculus, have a strong normative element.

An Illustration

As an illustration of the above points, consider the question of whether, for cohabiting couples who have not been previously married, the woman's becoming pregnant influences their 'time to marriage' (see e.g. Blossfeld and Rohwer, 1995; Blossfeld *et al.*, 1998). For a statistical analysis, it might be possible to find demographic data on potentially relevant covariates, such as the ages of the cohabiting partners, their social class, level of education, etc. Follow-up data from the beginning of the cohabitation status would then register the time to pregnancy, if any, and the time to marriage, or alternatively to the end of cohabitation or of follow-up.

In order to study this question, one obviously must first think about the influence of the measured covariates on the couple's 'rate of getting married'. If there were enough relevant demographic data one could either stratify according to these covariates or construct a sensible statistical model to account for their influence. In either case, it would be natural for an outsider to consider couples to be exchangeable if they share the same measured covariate values. Based on such data one could then estimate 'marriage rates' for groups of couples with shared covariate values. In particular, one could compare such rates before and after pregnancy was established, as functions of the time which the partners had spent living together. Empirical results from Blossfeld *et al.* (1998) show that the marriage rate starts to increase soon after pregnancy is established but then decreases, and, at the time the child is born, it is already close to the marriage rate of couples with similar covariate history who had not conceived a child.

Suppose for simplicity that there is so much data that a frequentist and a Bayesian analysis would have produced essentially the same marriage rate curves. From a frequentist point of view, their values would be interpreted as marriage frequencies averaged over a population, and from a subjectivist perspective they could be said to consist of momentary predictions concerning the marriage of 'a generic couple of unknown intentions' belonging to the exchangeable group of couples which is specified in terms of their shared covariate values. The definition of such a group would obviously depend on what covariate information were provided in the data.

The question now is what causal conclusions can be drawn from such a result? A straightforward claim in this situation would be to say that the observed differences in the marriage rates before and after pregnancy were actually evidence that 'pregnancies are causing marriages'. Such a conclusion does not seem to be completely warranted, however, because the observed rates do not distinguish important differences between couples. Because of effective contraceptive methods unplanned pregnancies for couples living in a stable relationship are rare, whereas most couples of a fertile age who want a child will have one without having to wait for too long. Therefore one might conclude that both pregnancy and marriage are typically pre-meditated events for which plans have existed in the minds of these people well before either conception or marriage. Considering pregnancy as a cause of marriage would necessitate comparing the two options 'pregnancy' and 'no pregnancy', but keeping the identity of the couple, including their plans and intentions, fixed. Conceptually, this would perhaps correspond to thinking of a situation in which the contraceptive method that was used had failed, resulting in an unplanned pregnancy, and in which abortion would not be an option. Such a comparison, however, is impossible in a study based on only demographic data, because the plans and

intentions of the couples remain unknown, and therefore they become potential confounders of the causal analysis. A completely plausible explanation of the observed data, without any true causal mechanism, is that the computed marriage rates before and after pregnancy are merely a reflection of a selection process taking place over time: those who intended to marry did so either already before pregnancy was established or soon thereafter, and were then 'selected away from the risk set', while those who did not have such an intention also did not marry, which explains the downward trend in the observed marriage rate.

This, of course, is not to say that, in some cases, pregnancy could really come as a surprise and then act as the direct cause of a decision to marry. In former times, such a sequence of events must have been quite frequent. Today, however, a causal claim would seem better justified if one were considering, for example, the issue of whether remaining childless could be considered as a cause of divorce. The basis of such a claim could be a similar comparison as above, but now between divorce rates before and after the birth of the first child (assuming that the couple did not have a child already when they married). From a causal perspective, the important difference between this and the previous question is that here it would not be unnatural to assume that a large majority of all couples entering marriage would then want a child. This hypothesis (which of course cannot be checked from the data either), if true, would effectively rule out the possibility that the contemplated cause, the birth of a child, or the lack of it, was in fact to some degree determined by an earlier unobserved confounding variable. A similar conclusion can be drawn even more strongly in the case, already mentioned above, of the study involving the effect of a loss of a spouse or a child in a car crash (Lehman *et al.*, 1987; Rosenbaum, 1999).

References

Arjas, E. and Eerola, M. (1993) On predictive causality in longitudinal studies. *Journal of Statistical Planning and Inference*, **34**, 361–386.

Blossfeld, H.-P. and Rohwer, G. (1995) *Techniques of Event History Modeling: New Approaches to Causal Analysis*. Erlbaum, Mahwah, NJ.

Blossfeld, H.-P., Hamerle, A. and Mayer, K.U. (1989) *Event History Analysis*. Erlbaum, Hillsdale, NJ.

Blossfeld, H.-P, Klijzing, E., Pohl, K. and Rohwer, G. (1998) Why do cohabiting couples marry? an example of causal event history analysis to interdependent systems. Manuscript.

Eerola, M. (1994) *Probabilistic Causality in Longitudinal Studies. Lecture Notes in Statistics*, **92**, Springer, Berlin.

Freedman, D. (1999) From association into causation: some remarks on the history of statistics. *Statistical Science*, **14**, 243–258.

Lehman, D., Wortman, C. and Williams, D. (1987) Long-term effects of losing a spouse or a child in a motor vehicle crash. *Journal of Personality and Social Psychology*, **52**, 218–231.

Mackie, J.L. (1965) Causes and conditions. *American Philosophical Quarterly*, **4**, 245–264.

Martikainen, P. and Valkonen, T. (1998) Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. *Lancet*, **348**, 208–213.

Parner, J. and Arjas, E. (1999) Causal reasoning from longitudinal data. Research Report A27, Rolf Nevanlinna Institute, University of Helsinki.

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

Robins, J. (1997) Causal inference from complex longitudinal data. In Berkane, M. (ed.) Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics *120*. Springer, New York, pp. 69–117.

Rosenbaum, P. (1999) Choice as an alternative control in observational studies. *Statistical Science*, **14**, 259–304.

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

Savage, J. (1972) *The Foundations of Statistics*. Dover, New York.

Author's Address

Elja Arjas, Rolf Nevanlinna Institute, FIN-00014, University of Helsinki, Finland.

Manuscript received: July 2000.