Causal Conjecture

Glenn Shafer[1]

Causal relations are regularities in the way Nature's predictions change. Since we usually do not stand in Nature's shoes, we usually do not observe these dynamic regularities directly. But we sometimes observe statistical regularities that are most easily explained by hypothesizing such dynamic regularities. In this chapter, I illustrate this process of causal conjecture with a few simple examples.

I first consider a negative causal relation: *causal uncorrelatedness*. Two variables are causally uncorrelated if there are no steps in Nature's event tree that change them both in expected value. They have, in this sense, no common causes. This implies, as we shall see, that the two variables are uncorrelated in the classical sense in every situation in the tree. When we observe that variables are uncorrelated in many different situations, then we may conjecture that this is due to their being causally uncorrelated.

I will also discuss three causal relations of a positive character. These relations assert, each in a different way, that the causes (steps in Nature's tree) that affect a certain variable X also affect another variable Y. This implies regularities in certain classical statistical predictions. The first causal relation, which I call *linear sign*, implies regularity in linear regression. The second, *scored sign*, implies regularity in conditional

---

[1] Faculty of Management, Rutgers University, Newark, New Jersey 07102, USA, and Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 OEX.

expectation. The third, *tracking*, implies regularity in conditional probability.

Chapters 5 through 10 of *The Art of Causal Conjecture* give a detailed account of sign and tracking, with an emphasis on their relation with causal thinking in philosophy. Here I will pay more attention to their relation with causal thinking in econometrics.

The main message of this paper is that Nature's probability tree is an adequate framework for causal inference; we do not always need more rigid frameworks such as stochastic processes or the Neyman-Rubin-Holland model. Another theme is that there are many interesting causal relations among variables, no one of which is so exclusively important as to merit the name "cause." The relations that I here call "X is a linear sign of Y" and "X tracks Y" are both important, and they are quite different; neither implies the other. They make precise very different ways previous authors have used the vague phrase "X causes Y."

1 Variables in a Probability Tree

Before plunging into a discussion of statistics, let us make sure that we have a common understanding of basic ideas: event, variable, probability, and expected value.

In classical probability theory, a random variable is a real-valued function on the sample space. In a probability tree, the sample space (the set of all ways the overall experiment can come out) is the set of paths down the tree, and thus a variable is simply a number determined by the path down the tree events take.

An event, in a probability tree as in the classical theory, is a subset of the sample space. I sometimes call such events *Moivrean events*, in order to distinguish them from the instantaneous events discussed in §1.1. (See Chapter 2 of *The Art of Causal Conjecture*.) The probability of a Moivrean event A is the same as the expected value of the variable that takes the value 1 on paths in A and the value 0 on paths not in A.

It is natural, when we are thinking about a probability tree, to say that we are concerned not merely with one probability measure, but with many—one for every situation. So when we speak of probability, expected value, variance, or covariance, we need to specify the situation S: $P_S(X=x)$ is the probability in S that X will eventually take the value x, $E_S(X)$ is the expected value of X in S, etc. This is illustrated in Figure 1.

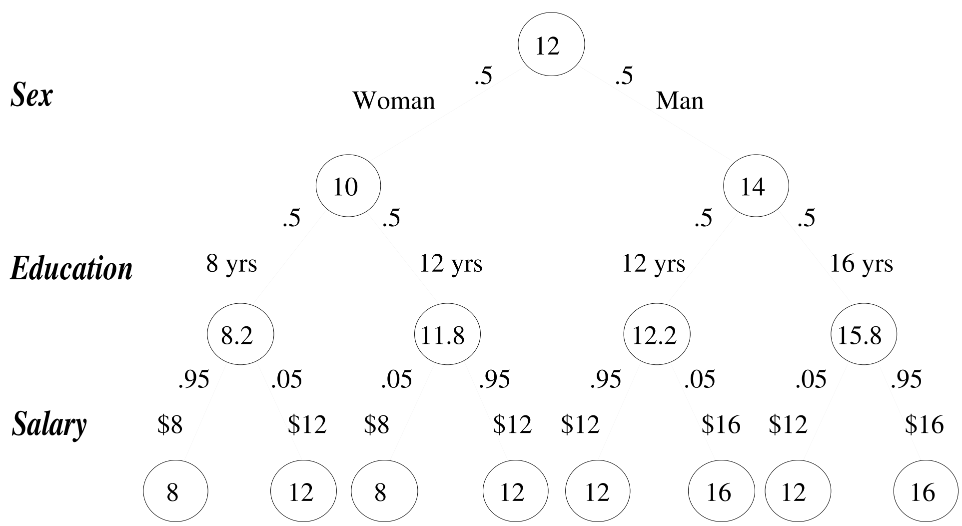

Figure 1 A probability tree for education and salary in an imaginary discrimatory society. This society educates men more than women, but there is some overlap. People are usually paid in proportion to their education, but employers may deviate from proportionality for an exceptionally capable or hapless employee, provided they stay within the range of pay customary for the employee's sex. The expected value of salary is shown in each situation (circle).

2 Causal Uncorrelatedness

As I have already said, two variables X and Y are *causally uncorrelated* if there are no steps in Nature's tree where they both change in expected value. Figure 2 shows a simple example and illustrates the theorem already mentioned: two causally uncorrelated variables X and Y are uncorrelated in the classical sense in every situation:

$$E_S(XY) = E_S(X)E_S(Y)$$

for all S. This theorem is easily proven by induction up the tree; see Chapter 8 of *The Art of Causal Conjecture*.

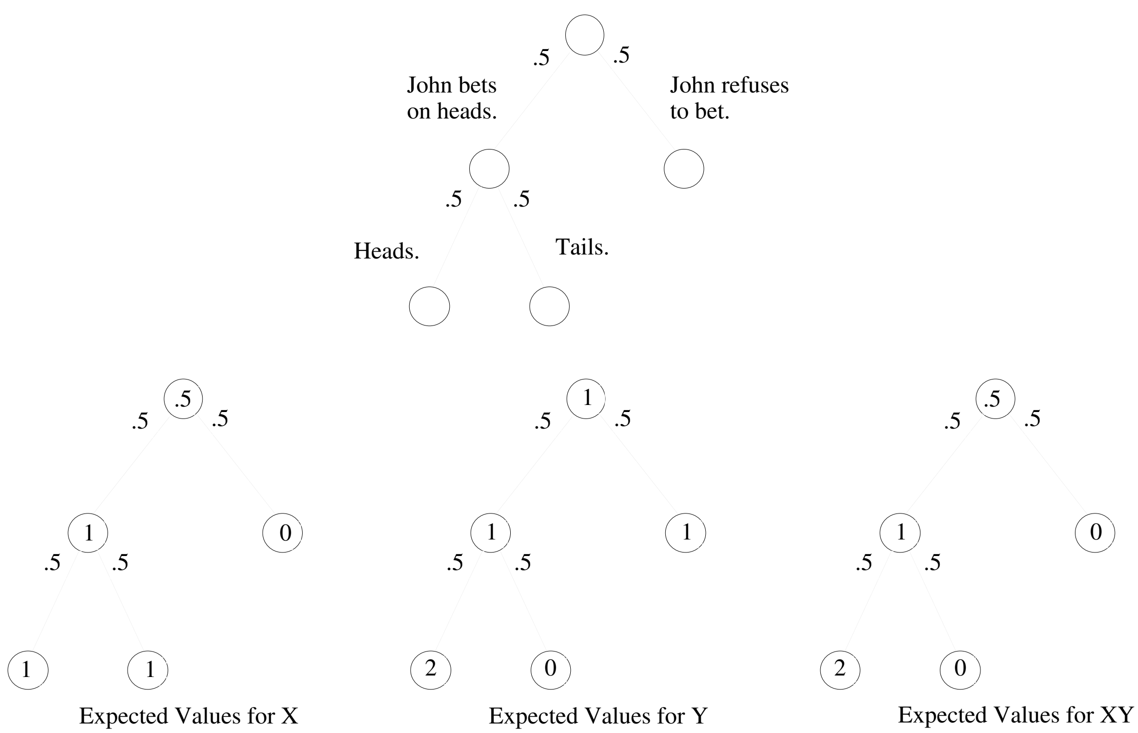

Figure 2 John, who has $1, decides whether or not to bet it on a fair toss of a coin. We write X for whether he bets (X=1 if he bets; X=0 if he does not), and we write Y for the number of dollars he has in the end. Then X and Y are causally uncorrelated; X changes in expected value only on the two steps at the top of the tree, while Y changes in expected value only at the two steps at the bottom. A glance at the expected values at the bottom of the tree confirms the consequent classical uncorrelatedness: E(XY) = E(X)E(Y) in each situation.

For some purposes, we may wish to strengthen causal uncorrelatedness to *causal independence*: X and Y do not change in probability on any step on Nature's tree. Causal independence implies classical independence in every situation: $P_S(X=x\&Y=y) = P_S(X=x)P_S(Y=y)$ for all S. Causal uncorrelatedness is more fundamental than causal independence, however, for it may hold even if the regularities Nature observes fall short of determining a complete joint probability distribution for X and Y (see Chapter 12 of *The Art of Causal Conjecture*).

3 Three Positive Causal Relations

Now we define the three positive causal relations studied in the remainder of the chapter (linear sign, scored sign, and tracking) and explain how they are related to classical prediction.

Classical statistical prediction considers a known joint probability distribution for random variables X and Y and supposes that we want to use this distribution to guess (or "predict") the value of Y from an observed value x for X. Of the many ways of doing so, three are most prominent:

- In order to predict Y linearly from x with the least mean squared error, we use the *linear regression:*

$$\hat{Y}(x) = E(X) + \frac{Cov(X,Y)}{Var(X)} [x-E(X)], \qquad \text{Equation 1}$$

  where E(X) is the expected value of X, Var(X) is its variance, and Cov(X,Y) is the covariance of X and Y. We call Cov(X,Y)/Var(X) the *regression coefficient* of Y on X.
- If we want to minimize our mean squared error in predicting Y from x, and we do not insist on the prediction being linear in x, we use conditional expected value:

$$\hat{Y}(x) = E(Y|X=x). \qquad \text{Equation 2}$$

This may or may not be linear in x. If it is linear in x, then it is necessarily the same as the prediction in Equation 1. If it is not linear in x, then it may be called the *nonlinear regression* of Y on X.

- Finally, if we want probabilities for Y rather than a single number as a guess for its value, we use the conditional probabilities:

$$P(Y=y|X=x). \qquad \text{Equation 3}$$

In practice, we often only hypothesize a joint distribution for X and Y and do not pretend to know this distribution exactly. In this case, we estimate the predictions in Equations 1-3 from data, and the choice among the three modes of prediction then depends in part on the amount of data. Linear regression is least informative but also requires the least data. Probability prediction is most informative but requires the most data.

As we have already noted, the variables in Nature's probability tree change in probability and expected value as Nature moves through the tree. Systematic relations between the changes for one variable and the changes for another, since they are dynamic aspects of Nature's tree, can be called causal relations between the variables. Here are three such relations:

- X is a *linear sign* of Y if on any step in Nature's tree where the expected value of X changes, the expected value of Y changes proportionally. In other words, there is a constant b such that

$$E_T(Y) - E_S(Y) = b\{E_T(X) - E_S(X)\} \qquad \text{Equation 4}$$

whenever T is a daughter of S and $E_T(X) \neq E_S(X)$. We call b the *sign coefficient*.

- X is a *scored sign* of Y if there is a function g such that on any step in Nature's tree where the probability distribution of X changes, the expected value of Y is equal to the change in the expected value of g(X). In other words,

  $$E_T(Y) - E_S(Y) = E_T(g(X)) - E_S(g(X)) \qquad \text{Equation 5}$$

  whenever T is a daughter of S and there is some value x for X such that $P_T(X=x) \neq P_S(X=x)$. We call g the *score function*.

- X *tracks* Y if for every function h of Y there is a function g_h such that the change in the expected value of h(Y) is always equal to the change in the expected value of $g_h(X)$ until after the value of X is settled. In other words,

  $$E_T(h(Y)) - E_S(h(Y)) = E_T(g_h(X)) - E_S(g_h(X)) \qquad \text{Equation 6}$$

  whenever T is a daughter of S and the value of X is not yet settled in S. We call g_h the *tracking function* for h.

The relation of tracking is obviously stronger than scored sign; if X tracks Y, then X is a scored sign of Y. There is no such implication between linear sign and the other two relations. Linear sign is stronger that the other two relations insofar as it insists on linearity, but weaker inasmuch as it imposes no condition on the change in the expected value of Y on steps where X does not change in expected value.

Our three causal relations (linear sign, scored sign, and tracking) correspond to our three classical concepts of prediction (linear regression, conditional expectation, and conditional probability) in the following way:

- If X is a linear sign of Y with sign coefficient b, then b is always the regression coefficient of Y on X . More precisely, for every situation S such that $Var_S(X) > 0$, there is a constant a_S such that the linear regression of Y on X is $a_S + bx$.

- If X is a scored sign of Y, with score function g, then g(x) is always equal, up to an additive constant, to the expected value of Y given X=x. For every situation S there is a number a_S such that

$$E_S(Y|X=x) = a_S + g(x)$$

whenever x is a value of X such that $P_S(X=x) > 0$.

- If X tracks Y, then for every x and y there is a number $p_x(y)$ such that

$$P_S(Y=y|X=x) = p_x(y)$$

for every situation S such that (i) $P_S(X=x) > 0$ and (ii) the value of X is not yet settled in S's mother. (Considered as a function of x for fixed y, $p_x(y)$ is equal to the tracking function g_h, where h is the function that assigns 1 to y and 0 to the other possible values of Y.)

In other words, each causal relation implies a certain stability in the corresponding classical prediction. Such stability can therefore provide evidence for causal conjecture: if the stability is observed statistically, then we may conjecture that the corresponding causal relation holds.

These causal interpretations reveal that the utility of linear regression, as opposed to that of the stronger forms of classical prediction, does not stem solely from the fact that we can estimate it with less data. Since its causal interpretation requires less regularity in nature, linear regression may sometimes have a valid causal interpretation even though the stronger forms of prediction do not, regardless of the amount of data available.

We should again emphasize that we discuss these three causal relations only because of their simplicity and their relation to familiar concepts of statistical prediction. There are many other causal relations, equally or more important for particular problems. Of particular importance are causal relations whose parameters can be estimated from experiments where treatments are assigned at random; these are discussed in Chapter 14 of *The Art of Causal Conjecture*.

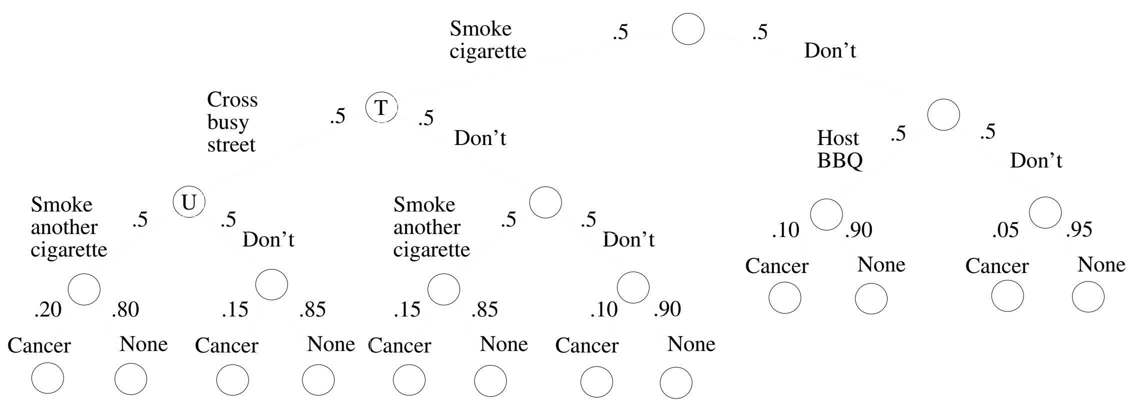

Figure 3 Nature's probability tree for Sally's smoking and cancer. For simplicity, this tree uses a very simple rule for the probability of cancer. At the last step on each path down the tree, the probability that Sally will get cancer is given by .05 + .05W, where W is the number of cigarettes Sally smokes, augmented by one if she hosts a barbecue or crosses a busy street.

4 Linear Sign

Figure 3 provides a simple example of linear sign. In this example, Sally smokes 0, 1, or 2 cigarettes and ends up with cancer or not. If we write X for the number of cigarettes she smokes and Y for whether she gets cancer (Y=1 if she does; Y=0 if not), then X is a linear sign of Y, with sign coefficient b = .05. Indeed, when we calculate the expected value for both X and Y in every situation, as in Figure 4, we see that on every step where X changes in expected value, Y changes in the same direction, .05 times as much.

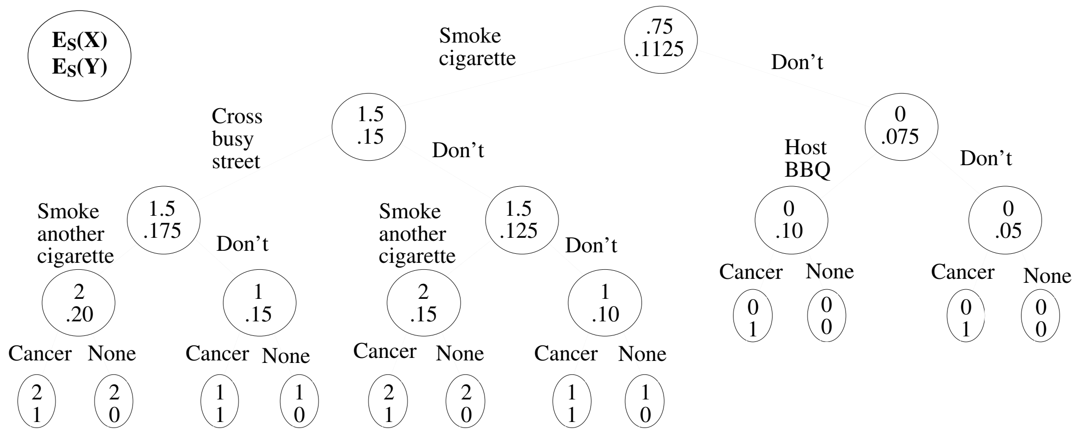

Figure 4 In each situation S, we have written the expected values $E_S(X)$ and $E_S(Y)$, where X is the total number of cigarettes Sally smokes and Y indicates whether she gets cancer. Since Y takes the value 1 if she gets cancer and 0 if not, $E_S(Y)$ is simply the probability she gets cancer.

The reader may use these expected values to verify that X is a linear sign of Y, with sign coefficient .05; on steps in the tree where E(X) changes, E(Y) changes .05 times as much. For example, on the step where Sally smokes her first cigarette, E(X) goes up by .75, and E(Y) goes up by .0375. On steps where E(X) does not change, the change in E(Y) is not required to be proportional to the zero change in E(X). On the step from T to U, for example, E(X) does not change but E(Y) goes up by .025.

As the tree is presented, E(X) goes up only on steps where a cigarette is smoked. This is not essential to the example; we can easily imagine a refinement that represents stressful events that increase the expected value of X without entailing the immediate smoking of a cigarette.

Linear sign is a causal relation between X and Y, not because it says "X causes Y," but because it relates causes of X to causes of Y. It says that the steps in the tree that are causes of X (in the sense that they change X's expected value) are also causes of Y, and that they affect Y b times as much. As we noted in the preceding section, this causal relation has a statistical implication. If X is a linear sign of Y, then the sign coefficient b is the regression coefficient of Y on X in every situation in which the variance of X is positive. (When the variance of X is zero, the regression coefficient is not defined.) Figure 5 illustrates this constancy. As we see there, only the regression coefficient is constant. The other part of the regression equation, the y-intercept, can change.

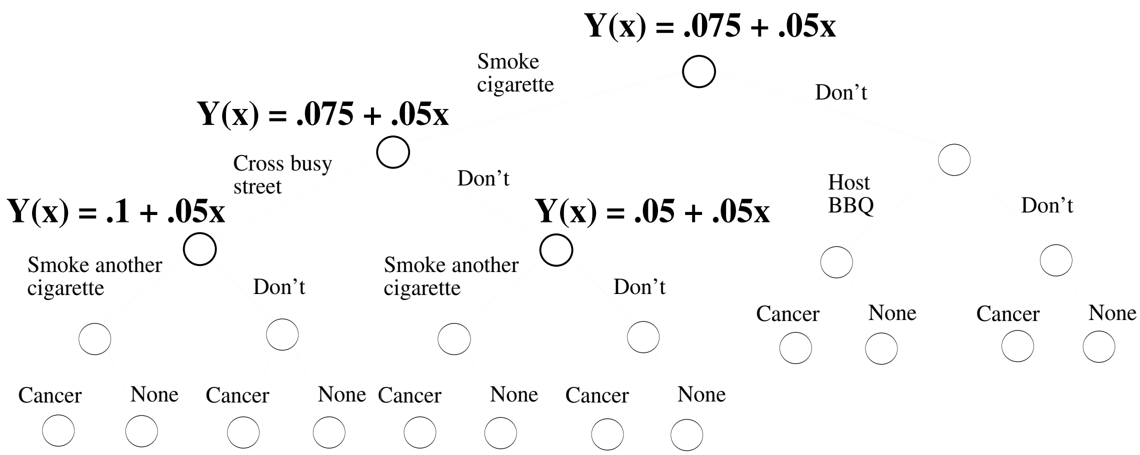

Figure 5 The regression of cancer (Y) on smoking (X) in the four situations in Figure 3 where it is defined. The regression coefficient is always .05, but the intercept of the regression changes. Crossing a busy street, for example, increases it from .075 to .1.

The condition that Equation 4 hold exactly for every step where E(X) changes is very strong, too strong to be satisfied very often. But the statistical implication of the condition is robust with respect to minor discrepancies. If the equation holds approximately on most steps where E(X) changes, then b will usually approximate the regression coefficient of Y on X. This follows from simple rules for the analysis of variance and covariance in probability trees. In general, the covariance between X and Y in a situation S can be decomposed over the steps below S:

$$\mathrm{Cov}_S(X,Y) = \sum_{\sigma} \Delta_\sigma X \cdot \Delta_\sigma Y \cdot P_S(\sigma) \,, \qquad \text{Equation 7}$$

where $\Delta_\sigma X$ is the change in the expected value of X on the step σ, and $P_S(\sigma)$ is the probability in S that the step σ will eventually be taken.[2] (In Figure 3, for example, if S is the initial situation and σ is step from T to U, then $\Delta_\sigma X$ is zero and $P_S(\sigma)$ is .5x.5 = .25.) When Y is equal to X, Equation 7 reduces to

$$\mathrm{Var}_S(X) = \sum_\sigma (\Delta_\sigma X)^2 \cdot P_S(\sigma) . \qquad \text{Equation 8}$$

We obtain a formula for b_S, the regression coefficient of Y on X in S, by dividing Equation 7 by Equation 8:

$$b_S = \frac{\mathrm{Cov}_S(X,Y)}{\mathrm{Var}_S(X)} = \frac{\sum_\sigma \Delta_\sigma X \cdot \Delta_\sigma Y \cdot P_S(\sigma)}{\sum_\sigma (\Delta_\sigma X)^2 \cdot P_S(\sigma)} . \qquad \text{Equation 9}$$

Since steps σ for which $\Delta_\sigma X$ is zero do not contribute to the sums in Equation 9, we can rewrite the equation as

$$b_S = \frac{\sum_\sigma (\Delta_\sigma Y / \Delta_\sigma X) \cdot (\Delta_\sigma X)^2 \cdot P_S(\sigma)}{\sum_\sigma (\Delta_\sigma X)^2 \cdot P_S(\sigma)} , \qquad \text{Equation 10}$$

where only σ with $\Delta_\sigma X$ not equal to zero are included in the sums. Equation 10 expresses b_S as a weighted average of the ratio $\Delta_\sigma Y/\Delta_\sigma X$ over steps below S where X changes in expected value. (The denominator is merely the sum of the weights.) When $\Delta_\sigma Y/\Delta_\sigma X$ is exactly constant over σ (always

[2] Equation 7 can be proven by induction from the bottom of the tree. See §3.5 of *The Art of Causal Conjecture* or §1 of Shafer (1997).

equal to b), the regression coefficient b_S is exactly constant over S (also always equal to b). More generally, when the ratio is approximately constant, the regression coefficient is approximately constant.

The asymmetry of linear sign should be emphasized. The demand that E(X) and E(Y) should change proportionally is symmetric in X and Y. But the condition that X is a linear sign of Y imposes this demand asymmetrically. It is imposed on all steps where E(X) changes, but not on all steps where E(Y) changes. As we have noted, there are steps in Figure 4 that violate Equation 4 inasmuch as they change E(Y) without changing E(X). The asymmetry of linear sign reflects the fact that the causes of X may not be the only causes of Y. Interleaved among the causes of the total number of cigarettes smoked (actual choices to smoke, or perhaps stresses that encourage these choices) are other causes (crossing the street, hosting a barbecue) of cancer.

One aspect of the asymmetry of linear sign is that it implies precedence: if X is a linear sign of Y, with a nonzero sign coefficient, then X precedes Y, in the sense that the value of X is settled either before the value of Y is settled or possibly at the same time. (The value of variable is settled in a given situation if and only if its expected value never changes below that situation. Thus the demand that Y change in expected value along with X forces Y to remain unsettled as long as X is unsettled.)

In order to make the concept of linear sign useful in practice, we must, of course, generalize it from the univariate to the multivariate case. The reader may consult Chapters 10 and 14 of *The Art of Causal Conjecture* for some steps in this direction.

To the best of my knowledge, the concept of linear sign has not been formulated by previous authors. The concept enters at an intuitive level, however, in many older discussions of causality, especially in explanations of diagrams where arrows are drawn between variables. See, for example, the discussion of such diagrams by the econometrician G. H. Orcutt (1952). In general, linear sign provides the most general and satisfying way of making precise the somewhat imprecise interpretation of linear path diagrams in econometrics, epidemiology, and the other social and biological sciences (see Chapter 15 of *The Art of Causal Conjecture*, Freedman 1991, and Koster 1996).

In the context of path diagrams, it is common to say that X is a cause of Y when there is a path from X to Y. As I have already argued, this is misleading. Linear sign is only one of many causal relations, and it is inappropriate in general to call a variable a cause. A variable is merely a global report on how events come out. Events in nature that change a variable's expected value are more appropriately called causes of that variable. If X is a linear sign of Y, then we should say not that X causes Y but rather that all the causes of X are also causes of Y.

5 Causal Uncorrelatedness Again

Although linear sign is a positive relation, it can be understood in terms of the negative relation of causal uncorrelatedness. Indeed, it follows directly from the definitions that X is a linear sign of Y with sign coefficient b if and only if X and Y-bX are causally uncorrelated. Thus X is a linear sign of Y if and only if

$$Y = bX + E, \qquad \text{Equation 11}$$

where X and E are causally uncorrelated.

Equation 11 reveals that the linear-sign interpretation of linear regression boils down to interpreting the uncorrelatedness of X and E as causal uncorrelatedness. Thus the linear-sign interpretation of linear path diagrams with uncorrelated errors (recursive systems of linear regressions) amounts to assuming that the uncorrelatedness is causal. This idea extends to a causal interpretation of systems of simultaneous linear equations; in this more general case, the equations do not represent linear-sign relations (indeed, they are no longer linear regressions), but the causal meaning of the equations is still expressed by assertions of causal uncorrelatedness.

This interpretation of simultaneous linear equations matches the intuitions underlying the classical treatment of the topic, in which variables are thought to be uncorrelated if they are affected by distinct events in nature. Consider for example, the simple model for demand and supply consisting of a demand equation,

$$Q = \alpha P + U_1, \qquad \text{Equation 12}$$

and a supply equation,

$$Q = \beta P + U_2. \qquad \text{Equation 13}$$

Here P is the price of a commodity, say wheat, and Q is the quantity sold in the market. The residual U_1, or $Q-\alpha P$, measures changes in the willingness of consumers to buy, while U_2, or $Q-\beta P$, measures changes in the willingness of producers to sell. Since there are events in nature that affect both U_1 and U_2, these variables are not causally uncorrelated. In fact, each of the four variables P, Q, U_1, and U_2 is causally correlated with each of the others, and this makes the model difficult to study; neither the elasticity of demand, α, nor the elasticity of supply, β, is identifiable. But as Jan Tinbergen suggested in 1930 (see Morgan, 1990, pp.180-182), if we can measure a variable, say rainfall R, whose causes affect $Q-\beta P$ but not $Q-\alpha P$, then we can replace Equation 13 with

$$Q = \beta P + \gamma R + U_2, \qquad \text{Equation 14}$$

where R is causally uncorrelated with U_1 and also with U_2, which now measures the effect of events that affect the demand curve without changing R. In the model given by Equations 12 and 14, α is identifiable.

6 Scored Sign

Roughly speaking, the concept of scored sign is generalization of the concept of linear sign, for it relaxes the requirement that the dependence of Y on X be linear. Instead of requiring that the change in the expected value of Y be the same as the change in the expected value of bX, we require that it be the same as the change in the expected value of some possibly nonlinear function g(X).

Scored sign is not strictly weaker than linear sign, however, for it requires slightly more than linear sign in another respect: it requires that the expected value of Y should not change on steps where the probabilities for X change but the expected value of g(X) does not. Linear sign imposes no condition on the possible change in expected value of Y on steps where X changes in probability but not in expected value. This point is illustrated by Figure 6, in which X is a linear but not a scored sign of Y. The example in Figure 6 may be atypical; the more common case is presumably that illustrated in Figures 3 and 4, where the linear sign is also a scored sign with a linear sign function.

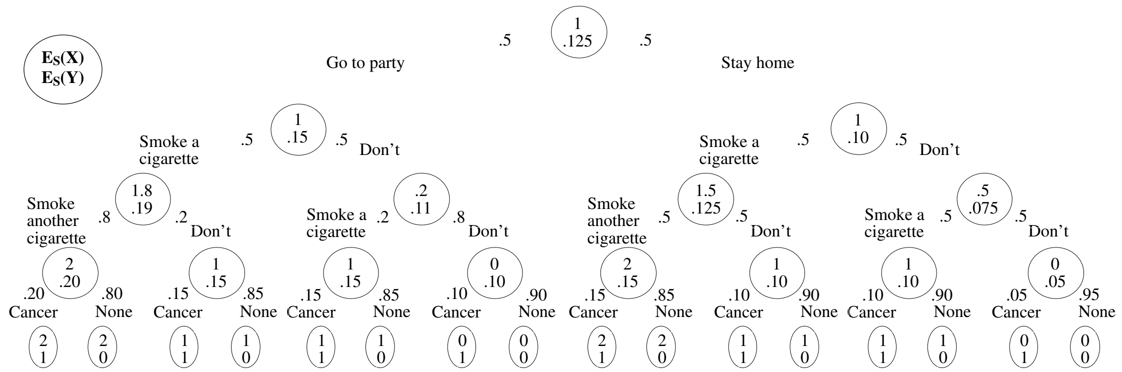

Figure 6 Here X, the total number of cigarettes Sally smokes, is once again a linear sign of Y, whether she gets cancer, with sign coefficient .05. But X is not a scored sign of Y. Although X does not change in expected value in the first step of the tree, it does change in probability there; the probabilities for 0, 1, and 2 cigarettes are .4, .2, and .4, respectively, if Sally goes to the party and .25, .5, and .25 if she stays home. Thus the fact that Y changes in expected value on this step is inconsistent with saying that X is a scored sign of Y with sign function .05X.

Even if not linear, the score function for a scored sign is often monotonic. For example, instead of the hypothesis that the number of cigarettes smoked is a linear sign of cancer, we might want to study the more realistic hypothesis that it is a scored sign of some index Y of lung disease, with a logarithmic score function, say g(x) = log(a + bx) for some real numbers a and b. But we can also consider score functions that are not monotonic. In this case, there may be quite substantial changes in Nature's expectations about X that are quite irrelevant to Y because they do not involve changes in the mean of g(X). This means in particular that X can be a scored sign of Y without preceding Y. Figure 7 gives an example.

The fact that X can be a scored sign of Y without preceding Y is another argument against the indiscriminate use of the verb "cause" to indicate any causal relation. Presumably we do not want to say that a cause comes after its effect.

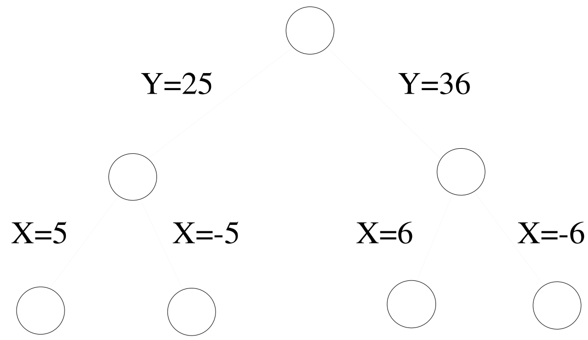

Figure 7 Here X is a scored sign of Y, with score function $g(x) = x^2$. In fact, $Y = X^2$. Notice that Y is always settled on the first step down the tree and X on the second; X does not precede Y.

7 Tracking

It can be shown that X tracks Y if and only if at the point where X is settled, its value is sufficient for predicting Y; nothing else that Nature knows helps further in the prediction. More precisely, if S and T are both situations where it is settled that X will equal x, then the probability distribution of Y in S is the same as in T.

As noted in §3, the statistical implication of X tracking Y is that until after X is settled, $P_S(Y=y|X=x)$ is the same in every situation S where it is defined. Here, (in contrast with the cases of linear sign and scored sign) the statistical implication is more than an implication; it is a characterization. Thus tracking, as opposed to linear sign and scored sign, can be characterized in a way that does not explicitly refer to changes from situation to situation.

The definitions given in §3 make it clear that tracking is strictly stronger than scored sign; if X tracks Y, then X is a scored sign of Y, with score function equal to the tracking function for the identity. The example of "X is a scored sign of Y" in Figure 7 is also an example of "X tracks Y." But in general, tracking is more exigent than scored or linear sign, inasmuch as it demands that X measure, in some sense, all the causes of Y up to the point where X is settled. Linear and scored sign, in contrast, allow other causes to be interleaved in the determination of Y. Since tracking is so much more exigent than sign, we may anticipate that it be less often realistic as a hypothesis.

Although X follows and tracks Y in Figure 7, there are many other examples where X precedes and tracks Y. A familiar one is given in Figure 8. Here is X is the number of heads in the first six spins of a fair coin, while Y is the number in the first eight spins. The reader will see immediately how to generalize this example; if we write X_n for the number of heads in the first n spins of a fair coin, then $X_1,X_2,...$ is a Markov process, and X_i precedes and tracks X_j whenever $i < j$.

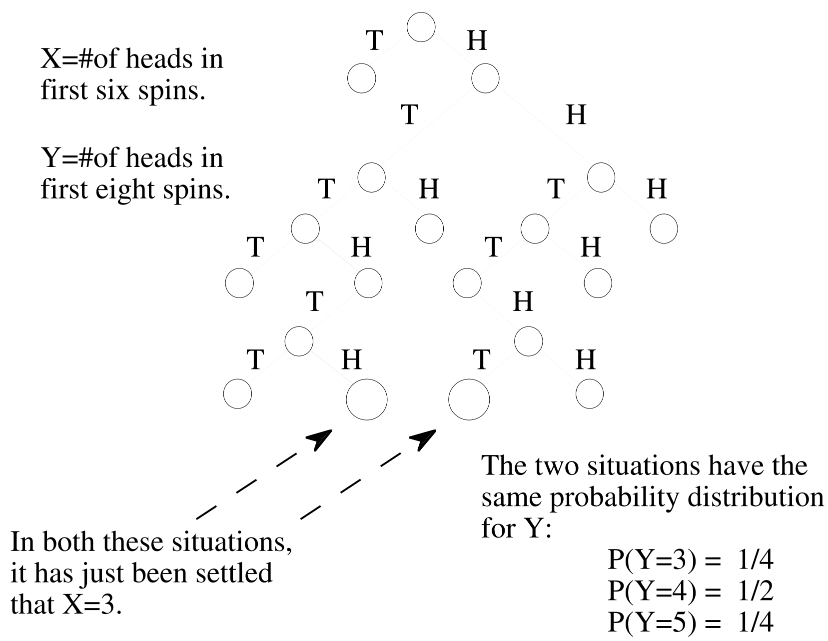

Figure 8 Here X precedes and tracks Y. At the point where the sixth spin has just been completed, the probability of a given number of heads in the first eight spins depends only on the number of heads obtained so far. Nothing further that Nature might have observed, including the order of the heads and tails so far, matters.

The example just given illustrates the idea of a stochastic process *unfolding in nature*. We may say that a stochastic process $X_1,X_2,...$ unfolds in nature if Nature observes the X_n in order and the conditional probabilities

$$P(X_{n+1}=x_{n+1}|X_1=x_1,...,X_n=x_n)$$

are Nature's probabilities for X_n at the point where she has observed $X_1=x_1,...,X_n=x_n$. This can be expressed in the language I have already developed in two equivalent ways:

- The event tree constructed by branching first on X_1, then on X_2, etc., is a simplification of Nature's tree.
- X_n precedes X_{n+1} in Nature's tree, and $X_1,...,X_n$ together track X_{n+1}.

It is what people usually have in mind when they use stochastic processes to model natural phenomena, although they seldom articulate it clearly, since the idea of Nature's tree is not yet part of the common wisdom.

Authors who use stochastic processes as causal models sometimes impose the condition that the process be Markov (see, for example, Arntzenius 1990 and Spohn 1990). But as Aalen (1990) has pointed out, the Markovian assumption is irrelevant to whether a stochastic process has a causal interpretation. The essential question is not whether $X_1,...,X_{n-1}$ can improve the prediction of X_{n+1} based on X_n, but whether anything else Nature has observed can improve the prediction of X_{n+1} based on $X_1,...,X_n$.

Because the idea of a stochastic process unfolding nature is not always well understood, let us pause to note that it is not enough, in order for this condition to be fulfilled, that Nature should observe the X_n in order and that their joint distribution should be verified by experience. This is illustrated by the fact that if the probability tree at the top of Figure 9 gives Nature's probabilities, then the probability tree at the bottom does not, and hence the process X_1,X_2 does not unfold in nature. If Nature has the tree at the top, then she does observe X_1 before X_2, and .5 is indeed the correct value for the conditional probability $P(X_2=0|X_1=0)$, but Nature does not have the probability .5 for $X_2=0$ when she observes $X_1=0$, because at that point she also observes something further that enables her to predict X_2 better.

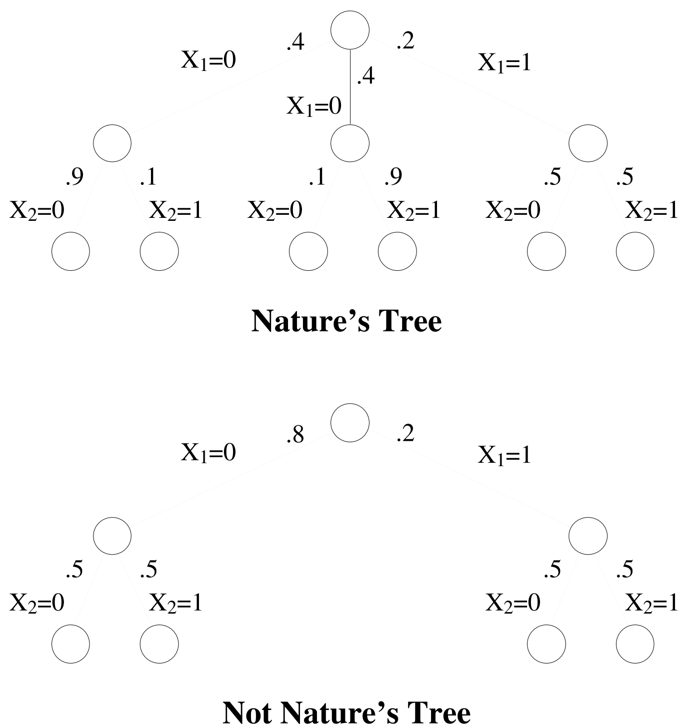

Figure 9 If the tree at the top gives Nature's probabilities, then the tree at the bottom does not, even though it gives the same order for observing X_1 ands X_2 and the same initial joint distribution for the two.

In recent years, the idea of a stochastic process unfolding in nature has been widely used in econometrics. Its proponents have not talked, as I do here, about Nature's probability tree, but they exploit a similar intuition. Clive W. J. Granger (1980), for example, begins his discussion of causality by imagining a time series which includes, at each point in time n, “all the knowledge in the universe available at that time.”

Once one has accepted a stochastic process as a causal model—i.e., once one has accepted that it unfolds in nature, further discussion of causality in econometrics is often conducted in terms of “Granger causation.” Suppose $Z_1, Z_2, \ldots$ unfolds in nature, and suppose $Z_n = (X_n, Y_n)$. Then the process $Y_1, Y_2, \ldots$ is said to *Granger cause* the process $X_1, X_2, \ldots$ if

$$P(X_n|X_1,\ldots,X_{n-1}) \neq P(X_n|Z_1,\ldots,Z_{n-1}).$$ [3]

This is more often expressed negatively; $Y_1, Y_2, \ldots$ is said *not to Granger cause* $X_1, X_2, \ldots$ if

$$P(X_n|X_1,\ldots,X_{n-1}) = P(X_n|Z_1,\ldots,Z_{n-1})$$

—i.e., if taking past Ys into account does not improve the prediction of the next X from earlier Xs.[4] Since the $P(X_n|Z_1,\ldots,Z_{n-1})$ are Nature’s probabilities (by the assumption that $Z_1, Z_2, \ldots$ unfolds in nature), $Y_1, Y_2, \ldots$ does not Granger cause $X_1, X_2, \ldots$ if and only if the $P(X_n|X_1,\ldots,X_{n-1})$ are Nature’s probabilities—i.e., if and only if $X_1,\ldots,X_{n-1}$ tracks X_n. Thus the assertion of Granger causation is simply the denial of an assertion about tracking in the context of an initial assumption about tracking.

Those who are already familiar with Granger causation will derive no new insight from its translation into the language of tracking. The translation, in itself, adds nothing to the concept. But it does open up a larger perspective. It places Granger causation in the context of a variety of other causal relations, thus suggesting new ways in which it can be weakened and therefore made more widely applicable. It also helps us articulate the limitations of statistical tests based on Granger causality; in a

---

[3] Granger 1980, p. 330.

[4] See Bruneau (1996), Hendry (1995), Florens and Mouchart (1982), and Florens and Fougère (1996). For critiques, see Leamer (1985) and Darnell and Evans (1990).

nutshell, they usually test whether $X_1,...,X_{n-1}$ tracks X_n in the context of the assumption that $Z_1,...,Z_{n-1}$ tracks Z_n, whereas it is the latter assumption that is most problematic. Here, as in other statistical work, the apparent rigor of statistical testing can divert attention from the assumptions with greatest causal significance to those that can be tested most sharply.

The causal relations that can serve as alternatives to Granger causation are countless, but because of the historical importance of simultaneous equations models, we should call special attention to causal uncorrelatedness and linear sign. As we learned in §5, these concepts can be used to formulate an understanding of simultaneous equation models that is rigorous but flexible, inasmuch as it does not require us to measure all of a variable's causes up to a given point. Putting simultaneous equation models into the framework of Granger causation amounts to making stronger and less realistic assumptions.[5]

References

Aalen, Odd O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal* 177-190.

Arntzenius, Frank (1990). Physics and common causes. *Synthese* **82** 77-96.

Bruneau, Catherine (1996). Analyse économétrique de la causalité: un bilan de la littérature. *Rev. écon. pol.* **106** 324-353.

Darnell, Adrian C., and J. Lynne Evans (1990). *The Limits of Econometrics*. Hants, England: Edward Elgar.

[5] Because they allow the interleaving of other causes, linear sign and causal uncorrelatedness are also distinct from and more flexible than the concept of linear Granger causality formulated by Florens and Mouchart (1985).

Florens, Jean-Pierre, and M. Mouchart (1982). A note on noncausality. *Econometrica* **50** 583-591.

Florens, Jean-Pierre, and M. Mouchart (1985). A linear theory for noncausality. *Econometrica* **53** 157-175.

Florens, Jean-Pierre, and Denis Fougère (1996). Noncausality in continuous time. *Econometrica* **64** 1195-1212.

Freedman, David A., Statistical models and shoe leather (with discussion). *Sociological Methodology* **21** (1991) 291-3511.

Granger, Clive W. J. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control* **2**, 329-352.

Hendry, David F. (1995). *Dynamic Econometrics*. New York: Oxford University Press.

Hendry, David F., and Mary S. Morgan, editors (1995). *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Koster, Jan T. A. (1996). Markov properties of nonrecursive causal models. *Annals of Statistics* **24** 2148-2177.

Leamer, Edward E. (1985). Vector autoregression for causal inference? *Carnegie-Rochester Conference Series on Pubic Policy* **22** 255-304.

Morgan, Mary S. (1990). *The History of Econometric Ideas*. Cambridge: Cambridge University Press.

Orcutt, G. H. (1952). Actions, consequences, and causal relations. *Review of Economics and Statistics* **34** 309-313. Reprinted in Hendry and Morgan, pp. 546-551.

Shafer, Glenn (1997). Mathematical foundations for probability and causality, in *Mathematical Aspects of Artificial Intelligence*, edited by Frederick Hoffman. Providence, Rhode Island: American Mathematical Society (Symposia in Applied Mathematics).

Spohn, Wolfgang (1994). Direct and indirect causes. *Topoi* **9** 125-145.