THE ART OF CAUSAL CONJECTURE
Glenn Shafer

Chapter 1. Introduction

Probability and causality must live together because both involve contingency. When I assign a probability to a coin's falling heads, I am saying that how it will fall is contingent. When I say the wind caused a tree to topple, I am saying that the wind and the toppling were contingent. The tree might have remained standing had the wind been less severe.

Jacob Bernoulli, the inventor of mathematical probability, taught that contingency, and therefore probability, are subjective. All things that exist, he said, are necessary and certain in and of themselves. Things can be contingent and uncertain—or partially certain—only relative to our knowledge. Probability is a degree of subjective certainty.[1]

Some may question the certainty of things in and of themselves, but Bernoulli was surely right to characterize contingency and probability as subjective. They are not subjective in the narrowest modern sense of the word; they do not pertain merely to a person's state of mind or disposition. But they are subjective in the broad sense in which Bernoulli used the word. They depend on the vantage point from which a person sees the world. They describe the person's knowledge and ability to predict. When I say it is contingent whether a coin will fall heads, I am acknowledging that I cannot calculate how it will fall. When I give a probability for its falling heads, I am expressing precisely my limited ability to predict how it will fall.

This book begins with a full acceptance of the subjectivity, in Bernoulli's sense, of the language of probability. Probabilistic ideas derive from a story about the relation between an observer and the world. As features of the observer's knowledge of the world, the probabilities in the story say something about the observer as well as something about the world. But in order to do justice to the objective as well as the subjective aspects of probability, we must stand back and recognize the diversity of ways the story can be used. It can be used to describe the situation of an actual person, but it can also be used in many other ways. In particular, it can be used to describe how the world works.

When we explain the working of the world in terms of objective probabilities, we are placing in the role of observer in the probability story an imaginary observer—we may call her nature—whose knowledge and prescience surpasses that of any actual observer but is nonetheless limited. Nature falls far short of God's omniscience, for she represents a hypothesized limit on the predictive ability that can be achieved by actual observers, human or artificial. Like actual observers, she watches events unfold step by step. Sometimes she can predict what will happen next, but more often she can only give probabilities.

Assertions about causation are also best understood as assertions about contingency as seen by nature. Ever since David Hume asked whether causation is anything more

---

[1] Bernoulli lived from 1654 to 1705. He made the statements paraphrased here in Part IV of his famous treatise on probability, *Ars Conjectandi*, which was published posthumously in 1713.

than the perception of constant conjunction,[2] philosophers have sought to provide a purely objective account of causation—an account of causation *as it is in the objects*. By thinking of nature as an observer, we abandon this valiant but futile quest. We preserve the objective aspects of causation—for it is the real objective world that nature observes—while acknowledging the ultimately subjective nature of all contingency. When we speak of causes, we are not speaking of the world as God sees it. We are speaking of the world as God has arranged for us to see it. We are speaking of regularities nature sees—regularities we glimpse when we pull ourselves up to one of nature's vantage points. We are speaking about the structure of nature's certainties and probabilities.

Nature is an idealization that cannot be avoided in an account of causality, but it is dangerously misleading to think of nature as pure object. By thinking of nature as an observer, we keep within our sight the role of actual observers in defining nature as a limiting idealization, and we thereby keep in touch with the subjective aspects of nature and causality.

1.1. Probability Trees

A full understanding of probability and causality requires a language for talking about the structure of contingency—a language for talking about the step-by-step unfolding of events. This book develops such a language based on an old and simple yet general and flexible idea: the probability tree.

Figure 1.1.1 Will Dennis remember to practice his saxophone?

Figure 1.1.1 is a probability tree for whether Dennis, a twelve-year old boy, will remember to practice his saxophone before dinner on a summer afternoon. He is least likely to do so when his friend Alex comes to his house and the two boys then go to Alex's house, for then he will be far too engaged in Alex's games to think of his saxophone. He is most likely to do so when he stays at home by himself and reads.

We can make sense of the contingencies and probabilities in a probability tree only in reference to an observer who watches events as they unfold. The next step at each point is contingent inasmuch as the observer is uncertain what will happen next, and the probabilities express the observer's limited ability to predict what will happen next. If there is no human or robot following Dennis about, then the only observer who can have the probabilities shown in Figure 1.1.1 is the demigod we call nature. So we can think of this figure as a coarse fragment of nature's tree. Nature follows all the happenings in the world, on this and all other afternoons, in great detail, and yet at each point marked by a

[2] Hume lived from 1711 to 1776. He developed his theory of causation in Part III of Book I of *A Treatise of Human Nature*, published in 1739, and he restated it in *An Enquiry Concerning Human Understanding*, published in 1748.

circle in the figure, nature can do no better in predicting which branch Dennis will take than by betting at the odds given by the probabilities on the branches.

Once we accept a probability tree as a fragment of nature's tree, it becomes natural to use causal language to describe what happens as events unfold. If Alex comes over and the two boys go to his house, we say that Dennis forgot to practice because he went to Alex's house. If he stays home and watches television by himself, we say that watching television contributed to his forgetting to practice. In this book, we will occasionally use this familiar language of singular causation, but our goal is to develop a language for describing nature's tree in the case where Dennis and Alex, or similar children, interact in similar ways over the course of many afternoons.

Another probability tree is shown in Figure 1.1.2. It is an elaboration of a tree drawn by Christiaan Huygens in 1676. Huygens's tree, shown in the original in Appendix A, may be first probability tree ever drawn. Huygens drew it in the course of writing out the solution of a problem he had published some years earlier—the problem of the gambler's ruin. Two players A and B begin with an equal number of tokens, and they repeatedly play a game, with the understanding that the winner of each round takes one of the other player's tokens. The first player to hold all the tokens wins the game and takes the stakes. Huygens wanted to find each player's probability of winning.[3] In the figure, it is assumed that each player begins with two tokens, so that the winner is the first player to gain a two-point advantage, and it is assumed that the odds in each round are 9 to 5 in favor of B. In theory play could continue forever, as indicated by the dots at the bottom of the figure; the tree is infinitely deep. Eventually, however, either A or B will win gain a two-point advantage, and the odds are 81 to 25 in favor of B,[4] so that B's probability of winning is 81/106.

The probabilities on the branches of a probability tree, which we call *branching probabilities*, add to one for all the branches at a given point. To get the probability for a particular path through the tree, we multiply together the branching probabilities on this path. The probability, for example, that Alex will come, Dennis and Alex will go to Alex's house, and Dennis will forget to practice his saxophone is 20%: .5x.4x1 = .2. The probabilities for all the paths through the tree again add to one. In Figure 1.1.1, there are ten paths through the tree. In Figure 1.1.2, there are infinitely many, and some are infinitely long; but the infinite ones all have probability zero, and the probabilities for the finite ones add to one.

---

[3] Huygens said this differently; he said he wanted to find the value of each player's hope or expectation—the portion of the stakes that would be due the player if the game were interrupted. It was only after Jacob Bernoulli introduced the idea of mathematical probability in *Ars Conjectandi* that Huygens's methods became methods for finding "the probability of winning."

[4] See Appendix A and Section 1 of Chapter 16. In general, if the odds are a to b in favor of B on each round, and a player must obtain a k-point advantage in order to win, the odds in favor of B winning are a^k to b^k. Edwards (1983) and Kohli (1975) discuss the different ways in which Pascal, Fermat, Huygens, Bernoulli, and De Moivre obtained this answer.

Figure 1.1.2 An elaboration of Christiaan Huygens's probability tree.

In standard elementary probability theory, the set consisting of all possible outcomes of an experiment is called its *sample space*, and the elements of the sample space are assigned probabilities adding to one. The idea of a probability tree is consistent with the standard theory, for the branches at a given point constitute the sample space for the experiment performed at that point,[5] and the paths through the tree constitute the sample space for the experiment represented by the tree as a whole.

In most twentieth-century mathematical and philosophical discussions of probability, especially since the influential work of Kolmogorov (1933), the sample space has been taken as fundamental. A probability tree, as we have just seen, includes a sample space but adds further structure to it.[6] This book is based on the contention that this further structure is fundamental to the meaning of mathematical probability. The insights into causality that we will gain by using probability trees provide one, but by no means the only, argument for this contention. There are a number of even stronger arguments for it.

First there is a historical argument. As Anthony Edwards (1987) has shown, probability-tree structure was fundamental to the reasoning of Pascal, who invented the mathematical theory of games of chance popularized by Huygens and made into a theory of probability by Bernoulli. There is no extant record of Pascal drawing a probability tree, but Huygens's example shows that it was natural do so even in the seventeenth century. Probability-tree reasoning continued to be fundamental for the meaning of probability in the work of De Moivre and Bayes (Shafer 1982, 1985).

Second, as we will see in Chapter 4, the probability-tree framework allows us to unify, directly in the foundations of the subject, the subjective and objective aspects of probability. A probability tree is best interpreted as a story about an observer's ability to predict. The branching probabilities give odds at which the observer might bet on the outcomes of each experiment. She can count on breaking even at these odds on a long sequence of steps in the tree, and having no further knowledge or insight, she cannot pick

---

[5] It may seem odd to use the word "experiment" for a game or for the process by which two boys decide where to play, but we will find this archaic usage convenient. In adopting it, we are following the example of Jacob Bernoulli, who used the Latin *experimentum* with similar breadth of meaning in *Ars Conjectandi*. One advantage of "experiment" is that it evokes, among statisticians, the idea of a randomized experiment. A probability tree is meaningful only if relative to some observer (if only a hypothetical observer such as nature) the branching probabilities have the same meaning as randomization probabilities have for a statistician observing a randomized experiment.

[6] There are other ways of adding dynamic structure to the bare idea of a sample space. The most widely studied way is to index variables by a measure of time, thus obtaining what is called a stochastic process (see Appendix H and Sections 7 and 8 of Chapter 2). Stochastic processes are sometimes put into a foundational role in the philosophy of probability, and they provide a framework in which many aspects of causality can be understood (Spohn 1990, Eerola 1994). But the greater simplicity and generality of probability trees makes them far more useful for these purposes.

bets that will do better than break even. The story expresses the subjective aspect of probability, since it is about an observer who can reasonably call the tree's probabilities her degrees of belief. It also expresses the objective aspect of probability, since it is about the position of this observer in the world. As we will see in Chapter 4, the fact that she cannot do better than break even says something about the frequencies with which events happen.

Third, the additional structure provided by the probability tree helps us understand the diversity of the uses of probability. In the sample-space framework, probability begins as pure mathematics. In the finite case, we merely have a set of numbers adding to one; in the general case, we have an axiomatic theory built on top of measure theory. The merit of this approach, as Kolmogorov (1933) emphasized, is that all interpretations of the numbers or axioms are equally legitimate. This makes it appear that all applications of probability can be accommodated. But directly interpreting numbers means interpreting them as a measure of something, and hence we seem obliged, in every application of probability, to identify something (frequencies, beliefs, etc.) that our probability numbers measure. With probability trees, in contrast, we begin with a unified story—a story about the beliefs of an observer and the frequencies that the observer observes, and it is this story, not a set of numbers or axioms, that we use in applications. A story can be used more flexibly than a set of numbers or axioms. It can be used as a model, to be tested by its fit with observations, but it can also be used as a standard of comparison or as one side of an analogy. This point is discussed further in the next section and in Section 6 of Chapter 4.

Fourth, the understanding of probability and causality based on probability trees generalizes readily to *martingale trees*. Martingale trees include probability trees, which have probabilities on all their branches, decision trees, which have probabilities only for some branchings (Raiffa 1968), and yet more general forms of partially probabilized trees. They are helpful for understanding decision making, quantum mechanics, and many other applications of probability. We study them in Chapter 12.

Fifth, at an abstract level the probability-tree approach allows the essential ideas of probability to be expressed without the technicalities of measure theory. This is explored in Chapter 11.

1.2. Many Observers, Many Stances, Many Natures

Although this book is concerned primarily with causality and therefore emphasizes probability trees in which nature is the observer, a clear understanding of the book's ideas is possible only if we remember that there are many different uses of probability trees. These different uses put different individuals, real or hypothetical, in the role of observer. We ourselves may or may not play the role of observer, and we may do so more or less wholeheartedly.

Players of games of chance place themselves in the role of observer in a probability story by agreeing to precise rules of play and then using thoroughly shuffled cards or other carefully crafted randomizing devices. Scientists do so by using random numbers generated by a computer program—numbers that are perfectly predictable for someone with full access to the program but maximally unpredictable for the scientists when they deny themselves this access. These players and scientists are full-fledged observers because they have full knowledge of the tree they will go through, including knowledge of the branching probabilities, and no other knowledge that can help them predict the path they will take.

Often we are less full-fledged observers. Even if we deliberately construct a tree, just as players in a game of chance do, we may be unable to state branching probabilities that define our ability to predict what path we will take through it. In such a case, the story about an observer with probabilities is still useful if we believe that the world is arranged in such a way that it is coherent—and consistent with what we will learn—to imagine such an observer, for we can then use our observations as we go through the tree to estimate some of the probabilities, or some aspect of them. Indeed, we may have set up the experiment that will take us through the tree precisely with this end in mind. This is the case when we sample from an engineering process or a finite population in order to learn about its characteristics.

Another stance is that of the manager or artificial agent who conjectures a probability tree in order to reason about what to do next. In this case, the tree and the branching probabilities that make it a probability tree are not immutable elements of the agent's knowledge. They are only conjectures, to be revised, perhaps drastically, as events actually unfold.

Other stances arise when we use a probability tree as an analogy or a standard of comparison. Analogies to probability trees can be helpful in weighing evidence; even if the evidence is not numerical, one way to assess its strength numerically is to draw an analogy between our situation and that of a probability-tree observer (Shafer and Tversky 1985). Such analogies are only analogies, and they put us at a distance from the probability-tree observer, who is only hypothetical. We stand at an even greater distance from a hypothetical probability-tree observer when we use her as a standard of comparison, for then we are comparing her ability to predict not with our own ability but with that of a method we are testing; we reject the method's claim to predictive ability if it can do no better than we might expect from a probability-tree observer who by assumption does not have that ability (Shafer 1994).

When we speak of causality, our stance is usually that of someone aiming for a partial understanding of something far too complicated to observe or even represent in its full relevant complexity. How do species evolve or thunderstorms form? How do people come to have cancer, and how do peoples come to rebel against their rulers? No one will ever completely describe nature's tree for these phenomena, but much of science consists of conjecturing partial descriptions. Nature's ability to predict represents the outer limit of understanding to which science aspires.

As we insist on the multiplicity of probability trees and observers, we must also acknowledge that in some problems there may be more than one nature—more than one level of knowledge from which the unfolding of events can be watched. After Henri Poincaré explained how the objective probabilities of statistical mechanics arise from the ignorance of initial conditions (see his *Science et méthode*, 1908), physicists discovered a more fundamental level of probability in quantum mechanics. So we are now accustomed to imagining two natures in physics, one who watches events at the macroscopic level, and a more informed one who watches events at the quantum level. In many other fields on inquiry, there is not so clear a choice between different levels of knowledge at which nature can be placed, and concomitantly, there is sometimes less clarity about what level is intended when causality is debated.

1.3. Causal Relations as Relations in Nature's Tree

The steps nature takes are the causes of where we end up. Dennis forgot to practice because Alex came over and he and Dennis decided to go to his house. We can also say that steps in nature's tree cause the probabilities of events to change. Consider the event that Dennis will practice.[7] Its total probability is initially 35%, but it changes to 28% if Alex comes over and to 42% if he does not, and it continues to change as we move on through the tree. As soon as the boys get to Alex's house, it is settled that Dennis will not practice, but the probability of his practicing continues to change along the other paths until the end.

The basic causal relations in nature's tree are relations between steps and changes in probability; the step—or the experiment in which it is taken—is the cause, and the change in probability is the effect. From these basic causal relations, we can define derivative relations between events, relations that are often useful for stating relatively simple causal conjectures. Here are four such relations.

1. One event *precedes* a second if it is always settled at least as soon as the second, no matter how we move through the tree.
2. Two events are *independent* if there is no place in the tree where both have their probabilities altered.[8]
3. One event *tracks* a second if the probability of the second is the same in any two places in the tree where the first happens and the same in any two places in the tree where the first fails.[9]
4. One event is a *positive sign* of a second if the probability of the second goes up whenever the probability of the first goes up, and goes down whenever the probability of the first goes down.[10]

This is not an exhaustive list. In later chapters, we will explore other ways of using events to describe causal structure.

Figure 1.3.1 How the probabilities and expected value for Dennis's distance from home change. Initially the expected value is .6 blocks. This

---

[7] In order to speak of the probability of an event, we must think of the event as a subset of the sample space. The event that Dennis practices can be thought of in this way: it is the set consisting of the five paths through the tree that end with a step labeled "remember."

[8] More precisely, there is no branching in which both can have their probabilities altered. As we will see in Section 2 of Chapter 5, it is possible that one event might have its probabilities altered on certain branches, while another has its probabilities altered on other branches of the same branching. If this happens, but there are no branches anywhere in which the two events both have their probabilities altered, then we say the two events are *weakly* independent.

[9] The idea of an event happening or failing in a certain place in a tree is explained in Section 2 of Chapter 2.

[10] The probability of the second is allowed to change arbitrarily when the probability of the first does not change at all.

becomes 1.2 if nature steps to the left (Alex comes over) and 0 if nature steps to the right (Alex does not come over).

Probabilities for variables also change as nature steps through her tree. Consider the distance from home that Dennis will be during the afternoon. If he stays at home, he is 0 blocks away; if he goes to Sigmund's house, he is 1 block away; if he goes to Alex's house, he is 2 blocks away. Figure 1.3.1 shows how the probabilities and expected value[11] for this variable change. Its value is always settled before the end. It is settled, for example, as soon as Alex does not come over. It is also settled (to the same value, 0) when Dennis and Alex decide to stay at Dennis's house.

The causal relations we defined for events generalize readily to variables. In fact, they generalize in many different ways, because a step in nature's tree may affect a variable in many different ways. Here are a few of the possibilities.

1. One variable *precedes* a second if it is always settled at least as soon as the second, no matter how we move through the tree.
2a. Two variables are *independent* if there is no place in the tree where both have their probabilities altered.
2b. Two variables are *uncorrelated* if there is no place in the tree where both have their expected values altered.
3a. One variable *strongly tracks* a second if the probabilities of the second are the same in any two places in the tree where the first is settled to have the same value.
3b. One variable *tracks* a second *in mean* if the expected values of the second are the same in any two places in the tree where the first is settled to have the same value.
4a. One variable is a *positive sign* of a second if the expected value of the second goes up (down) whenever the expected of the first goes up (down).
4b. One variable is a *linear sign* of a second if whenever the first changes in expected value, the second changes proportionally.

These relations are all defined more precisely and studied in detail in later chapters.

We must emphasize the indirection with which these relations inform us about causation. They do not say that one event or variable causes another, for events and variables are not causes. They tell us instead about how things that affect or cause one event or variable (namely, steps in nature's tree or experiments that produce these steps) are placed relative those that affect or cause another. Here are restatements of the definitions for events that make this explicit.

1. One event *precedes* a second if the last step in the tree affecting the first always precedes or coincides with the last step affecting the second.
2. Two events are *independent* if no experiment in the tree influences both.

---

[11] The idea of the expected value of a variable is reviewed in Section 5 of Chapter 3 and in Section 4 of Appendix D.

3. One event *tracks* a second if until the first is settled, steps that affect the second do so only insofar as they affect the first.
4. One event is a *positive sign* of a second if steps that promote the first also promote the second, and steps that hinder the first also hinder the second.

It is important to state precisely the causal content of the assertion that two variables are independent. It is sometimes said that two variables are independent if neither causes the other. But this is misleading and confusing, for a variable is never a cause. Steps in nature's tree are causes. Two variables are independent if they have no common causes.

Similarly, we should avoid the temptation to use the word "cause" in the place of "track" or "sign." The temptation may be especially great in the case of "sign." If an increase or decrease in the probability of E is always matched by a similar change in the probability of F, why not say that E is a cause of F? This accords with the way some philosophers have written about probabilistic causation.[12] But once we understand that causes operate in nature's probability tree, not at the level of the sample space, we see that little is gained by this kind of talk. It adds nothing to our understanding of the sign relation, and the confusion it may engender can have undesired consequences. It may seem harmless to label the variable "number of cigarettes smoked" as a cause of illness; this label points clearly to real causes: individual acts of smoking. But it is scarcely harmless to label membership in a disadvantaged group as a cause of the disadvantage, for this only distracts attention from what may be seen as the true causes when we take a broader view: acts of discrimination against the group.

1.4. Evidence

Causal conjectures are assertions about nature's tree. What kinds of evidence can we find for such assertions? This is a large question, with a variety of answers.

In some cases, we do accompany nature through part of her tree—we observe relevant events unfold. It is generally agreed that this kind of evidence—longitudinal evidence—is the strongest we can obtain for causality. Sometimes it is gathered formally; in other cases, it is accumulated informally as we experience the working of the world.

Although longitudinal evidence is always to be preferred we can sometimes find useful evidence in more limited data. Consider the simple case in which the experiment represented by a certain probability tree is repeated many times, and we measure certain variables each time, so that we finally learn, to some degree of approximation, nature's probabilities for these variables at the beginning of the tree. Suppose, for example, that we are told on many different afternoons whether Dennis has practiced his saxophone and how far from home he is at dinner time. If we obtain this information for enough afternoons, we will eventually have good estimates of the probability that he practices, the probability that he ends up a certain distance from home, and even the joint probability of the two events. We may be able to obtain these probabilities without observing the path taken through the tree each afternoon and and without even learning what the tree looks like. How can we use the probabilities to make conjectures about the structure of the tree?

---

[12] See, for example, Humphreys (1989) and Spohn (1990).

We are now discussing the kind of information most often considered in statistical theory: probabilities or expected values estimated from observations from a sample space.[13] The question we are asking goes beyond standard statistical theory, however. We are asking how the static and therefore non-causal information carried by a sample space can be used to form conjectures about underlying dynamic and causal probability-tree relations.

In order to make further progress, let us focus on the particular casual relations defined in the preceding section: probability-tree independence, tracking, and sign. How can these relations be discerned in sample-space information? What is the evidence of them there? As it turns out, this question has some relatively simple answers. Here are three of them:

- If two events or variables are independent in the probability-tree sense of the preceding section (they have no common causes), then they are also independent in the sample-space sense (their probabilities multiply).
- If two variables are uncorrelated in the probability-tree sense of the preceding section, then they are also uncorrelated in the sample-space sense (their expected values multiply).
- If one variable is a positive sign of another, then the two variables are positively correlated in the usual sample-space sense.

Although statements of this type were not rigorously formulated before this book was written, variants of them have long been understood and used at an informal level. We all recognize that "causation implies correlation," and we often use correlation as evidence of causation. As Figure 1.4.1 indicates, we often conjecture causation from correlation.

Figure 1.4.1 Proof and conjecture.

Conjecture from correlation to causation is notoriously unreliable; hence the maxim, "Correlation does not prove causation." Any two variables are either uncorrelated, positively correlated, or negatively correlated in the sample-space sense, but it does not follow that they are uncorrelated in the probability-tree sense or that one is a positive or negative sign of the other. On the contrary, these probability-tree relations are quite rare. Most sample-space correlations, whether zero, positive, or negative, have no simple causal explanation. Yet correlation is evidence of causality, and when buttressed with other knowledge and evidence, longitudinal or experimental, it can provide the basis for causal conjecture.

It sometimes happens that we can observe or estimate more than one set of probabilities associated with a probability tree. Thus far we have been considering

---

[13] Sample-space theory is reviewed in Appendices D though H.

probabilities in the initial situation, at the beginning of the tree; it is with reference to these probabilities that we have said given variables will be correlated or uncorrelated. But as we learned in Section 3, probabilities for events and variables change as nature moves through her tree. The events and variables have well-defined probability distributions, in the sample-space sense, corresponding to every situation nature might be in as she moves through the tree. (In the case of Figure 1.1.1, for example, we can speak not only of the overall probability that Dennis will remember to practice, but also of the probability of that event when he arrives at Sigmund's house, or of its probability when he arrives at Alex's house, for these are also situations in the tree.[14] Similarly, in the immense fragment of nature's tree that tracks the development of lung cancer in the world's population, we speak not only of the overall probability of lung cancer, but also of probabilities in more specific situations—after a person's sex has been determined, after his or her country of residence has been determined, and so on.) As it turns out, the statements we made above can be strengthened to refer to these other situations as well as the initial situation:

- If two events or variables are independent in the probability-tree sense, then they are also independent in the sample-space sense in every situation.
- If two variables are uncorrelated in the probability-tree sense, then they are also uncorrelated in the sample-space sense in every situation.
- If one variable is a positive sign of another, then the two variables are positively correlated in the sample-space sense in every situation until after the first variable has been settled.

Since the logical implications (corresponding to the arrows to the right in Figure 1.4.1) yield more in this formulation, the reverse conjectures (the arrows to the left) can be better supported. As we observe that two variables are positively correlated in more different situations, it becomes less plausible that these correlations should all be accidental, and we have a stronger case for causality. If, for example, we observe that high rates of cigarette smoking are correlated with high rates of lung cancer in many different countries, for men and for women, for farmers and for coal miners, then we may well conjecture that the relation is causal—i.e., that the rate of smoking is a positive sign of lung cancer in the probability-tree sense.

The other probability-tree relations we discussed in Section 3 also have important implications for sample-space probabilities. Here, for example, is an implication that justifies using the stability of a regression coefficient[15] to support a causal interpretation:

- If X is a linear sign of Y, then the regression coefficient of Y on X is the same in every situation until after X is settled.

And here are implications that give substance to the idea that conditional independence and partial uncorrelatedness[16] can have causal meaning:

---

[14] The condition that Dennis stays at home does not, in contrast, correspond to a single situation in this probability tree. There are two situations where it becomes settled that he will stay at home.

[15] See Appendix E.

[16] Conditional independence, partial uncorrelatedness, and other analogous concepts are studied systematically in Appendix F.

- If X strongly tracks Z, and Y and Z are independent in the probability-tree sense (have no common causes) posterior to X (i.e., after X is settled), then Y and Z are conditionally independent given X, in the sample space sense, in every situation.
- If X is a linear sign of Z, and Y and Z are uncorrelated modulo X in mean (i.e., do not change in mean except when X changes in mean), then Y and Z are partially uncorrelated linearly accounting for X in every situation.

We began this section by assuming that the same probability tree is traversed many times. An equivalent assumption is that this probability tree appears as a subtree, with the same probabilities, many times along every path through nature's tree. When we put the matter this way, we see that we are really dealing not with a single pair of variables X and Y, say, but with successive variables $X_1,X_2,...,X_n$ and $Y_1,Y_2,...,Y_n$. Since the subtree where X_i and Y_i are played out looks the same no matter which path we take through the tree, we can say, using language familiar to mathematical statisticians, that the (X_i,Y_i) are independent and identically distributed—the pair (X_i,Y_i) is independent of the pair (X_j,Y_j) but has the same probabilities. And when we say that X is a linear sign of Y, say, we are really saying that X_i is a linear sign of Y_i for each i. Fortunately, the assumption that the successive variables are independent and identically distributed can be relaxed in most cases. If we are studying the relation of linear sign, for example, it is sufficient that these variables be uncorrelated and have appropriately related initial expected values. This remains, however, an important and substantive assumption. The most difficult part of statistical modeling is identifying variables—identifying units and attributes of the units that we can expect to be related to each other in stable ways. (See, e.g., Draper et al. 1993.)

1.5. Measuring the Average Effect of a Cause

The ideas of tracking and sign assume a high degree of uniformity across nature's tree. This is appropriate and useful when we have relatively little data, for then we are best advised to extrapolate the little we do know. But when we are able to make extensive observation, we usually find that nature is not uniform, and our task becomes the measurement of average effects rather than uniform effects. The shift from uniform to average effect is more profound than we might at first think, for it requires new ways of representing causes.

Figure 1.5.1 illustrates the complications that can arise when effects are not uniform. This figure tells the story of Mark, whose chances of marrying Peggy, his childhood sweetheart, depend partly on how far he goes in school. They improve if he goes to college and improve even more if he graduates. Suppose we ask, a bit obtusely, about the effect of his dropping out of college. It is evident that the effect is not uniform, and we cannot even speak of an average effect unless we somehow specify the point in time from which an average might be calculated. Considered at the point in time where he has just finished high school and might not even go to college, dropping out of college would raise his chances on average. Considered at the point where he has just decided to go to Notre Dame or NYU, it would lower his chances. Considered at some other point not shown in the tree—say the point at which he is writing Peggy about his boredom with his classes, it might have some other effect.

Figure 1.5.1 Will Mark marry Peggy?

Figure 1.5.2 Mark's dropping out of college, considered at the point where he has just finished high school. At this point, it raises his chances of marrying Peggy.

We have already emphasized that causes are steps in nature's tree, not events. Now we see another aspect of this point. It makes sense to speak of the average effect of an event only if we have implicitly or explicitly situated that event—only if we have somehow specified concomitant events whose happening situates it. In general, a cause is a collection of steps in the tree, and in order to specify such a collection of steps, we need to specify both their starting points (this is done by the concomitant events) and their ending points. Figures 15.5.2 and 15.5.3 apply this thought to the story about Mark. The first figure shows Mark's dropping out of college considered at the point where he has finished high school but not decided on college, while the second figure shows the same event considered at the point where he has just gone to college.

Figure 1.5.3 Mark's dropping out of college, considered at the point where he has just decided to go to college. At this point, it lowers his chances of marrying Peggy.

Our conclusion is that an event must be supplemented with concomitants in order to specify a cause. A similar conclusion applies to variables. It is well accepted in the statistical literature that concomitants can often improve the precision of the statistical measurement of the effects of an event or a variable. One of the main contributions of this book to statistical thinking is the conclusion that concomitants have a much more fundamental role—they are needed even in order to define what we mean when we call an event or variable a cause. The implications of this conclusion are explored in Chapter 14.

1.6. Causal Diagrams

Although careful statistical investigation of causality must usually fall back on the concept of average effect, it can be useful, when extensive statistical information is not available, to hypothesize causal relations for whole networks of variables in terms of approximately uniform effects. This is often done using diagrams.

One of the contributions of this book is to show how precise interpretations can be provided for causal diagrams. As it turns out, a causal diagram can often be interpreted in several different ways as a statement about nature's probability tree. This provides an opportunity to add depth to the causal claims made on behalf of a diagram. Instead of accepting or rejecting the vague claim that the diagram is causal, we can demand that this claim be made more specific.

From the sample-space point of view, the causal diagrams with which we are concerned say something about how variables can be predicted, in the sample-space sense, from other variables. There are two broad types of these prediction diagrams: *path diagrams* and *relevance diagrams*. Diagrams of both types consist of variable names connected by arrows. In path diagrams, the arrows have numbers attached to them, which are to be used as coefficients in linear prediction. In relevance diagrams, numbers are not given, but the choice of arrows is supposed to indicate which variables are most relevant to the prediction of other variables.

Figure 1.6.1 is a simple example of a path diagram. It indicates that the yield of wheat in a particular field can be predicted from the field's phosphate level and acidity, according to a linear equation,

$$\text{Yield} = \alpha + \beta_1\text{Phosphates} + \beta_2\text{Acidity} + \text{Error}, \qquad (1.6.1)$$

where α, β_1, and β_2 are numerical coefficients, perhaps unknown prior to statistical investigation, and Error represents the influence of other factors, which are uncorrelated[17] with phosphates and acidity. If statistical investigation shows the diagram and the formula to be statistically valid, then we may conjecture that it also has some causal validity. We may want to say, for example, that when a field's phosphate level changes by one unit, "other things being equal," its yield of wheat increases, on the average, by β_1 units.

Figure 1.6.1 A path diagram for the yield of wheat. The double-headed arrow joining Phosphates and Acidity indicates that these variables may be correlated in the sample-space sense.

One of the several possible probability-tree interpretations of this path diagram is provided by conjecturing that phosphate level and acidity level are together a linear sign

[17] Here we mean uncorrelated in the usual static sample-space sense, not uncorrelated in the dynamic probability-tree sense explained in Section 4. See Section 4 of Appendix D.

of yield, in the sense that on any step in the probability tree where one or both of them change in mean (i.e., expected value), yield changes in mean proportionally, with β_1 and β_2 as the constants of proportionality. If we write $\Delta\mathbf{E}$ for the change in expected value on a given step, then we can express this more precisely by saying that

$$\Delta\mathbf{E}(\text{Yield}) = \beta_1\Delta\mathbf{E}(\text{Phosphates}) + \beta_2\Delta\mathbf{E}(\text{Acidity}),$$

whenever $\Delta\mathbf{E}$(Phosphates) - 0 or $\Delta\mathbf{E}$(Acidity) - 0. Under this conjecture, (1.6.1) will represent the linear regression of yield on phosphates and acidity in every situation in the tree until after the phosphate level and acidity level are settled, possibly with different values of α in different situations, but always with the same values for β_1 and β_2. To support the conjecture, we might investigate a number of different situations, checking whether the regression coefficients are indeed constant across situations. Or, alternatively, we might investigate changes directly, by adding phosphates or lime. Either approach will require us to specify to some degree what kinds of changes enter into the probability tree we have in mind, and this will take us beyond the vague and often dangerous phrase, “other things being equal.”

Relevance diagrams, in contrast with path diagrams, do not involve numbers. The information they carry is all in the arrows, which encode conditional independence relations. The variables in the diagram are ordered in a way with the arrows (all the arrows point from an earlier to a later variable in the ordering), and each variable is independent, given its parents, of all its predecessors in the ordering. There are a number of different types of relevance diagrams, corresponding to different sample-space conditional independence relations. The most common types are Markov diagrams, which use conditional independence in the standard sense, and linear relevance diagrams, which use partial uncorrelatedness. Markov diagrams, supplemented with conditional probabilities for each variable given its parents, are widely used in artificial intelligence, where they are called Bayes nets (see Chapter 16). Linear relevance diagrams, which are more often used in the social sciences, can be thought of as path diagrams in which certain correlations are zero.

Figure 1.6.2 shows a typical linear relevance diagram. It was constructed by Spirtes, Glymour, and Scheines (1993), on the basis of data collected and first analyzed by Rodgers and Maranto (1989). It indicates that the frequency with which the work of an American university psychologist is cited can be predicted linearly from four variables: the psychologist’s general intellectual ability, as measured by a standardized test when the psychologist first enters college, the quality of the institution where the psychologist first works after graduate school, the total number of items the psychologist publishes, and the number completed while the psychologist is still in graduate school. The other variables—perceived quality of the program in which the psychologist earns the doctoral degree and whether the psychologist is a man or a woman—do not (markedly) improve the accuracy of the linear prediction once these four are used.

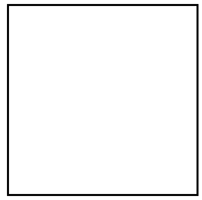

Figure 1.6.2 A relevance diagram for the frequency with which the publications of an academic psychologist are cited.

Figure 1.6.2 can be interpreted causally by interpreting each partial uncorrelatedness relation in the way suggested in Section 4; we conjecture that the parents of each variable in the diagram are linear signs of that variable, and that the variable is uncorrelated with earlier variables modulo its parents in mean. However, as we will see in Chapters 9, 10, and 15, this is only one of several possible probability-tree interpretations.

1.7. Humean Events

When we insist that an event is not a cause, we may be emphasizing too much one particular sense of the word "event." Following the established practice of probabilists and statisticians, we have been calling subsets of the sample space events. This sense of "event" was already prominent in Abraham De Moivre's book, *The Doctrine of Chances* (1718), and hence we may call it the *Moivrean* sense. But there is another sense of "event," more common in everyday language and often favored by philosophers. An event in this sense is something localized in time and space, perhaps a change in some object or circumstance at a particular time. We may call this the *Humean* sense of the word, since it is often used in discussions of Hume's analysis of causation.

The Moivrean and Humean senses of "event" are not always adequately distinguished in discussions of probability and causality, but they are readily distinguished in a probability tree. A Moivrean event is a set of complete paths through the tree, while a Humean event is a more localized collection of individual steps or chains of steps, as in Figures 1.5.2 and 1.5.3. In general, a Humean event can be specified in terms of its starting points and ending points, and it can be thought of as consisting of all the chains of steps from one of the starting points to one of the ending points (see Section 4 of Chapter 2). The Humean event in Figure 1.5.2 consists of three chains: the chain from R to U, the chain from R to V, and the chain from R to W. The Humean event in Figure 1.5.3 consists of individual steps on three different paths.

We can better appreciate the local nature of Humean as opposed to Moivrean events if we recognize that the meaning of an event includes a specification of alternatives. A Humean event is most local when it consists of a single step, and then the alternatives are other steps from the same starting point. The alternatives to a Moivrean event include all other paths through the tree, and these may be spread out across time and space. Figure 1.7.1 illustrates the point with reference to the event that Dennis and Alex go to Sigmund's house. As a Humean event, this is a single step: one of three things the boys might do after Alex comes over. The alternatives are that they stay at Dennis's house or go to Alex's house. As a Moivrean event, it consists of the two paths that go through this step, and its alternatives include all the other paths through the tree, including paths that begin with Alex not coming over.

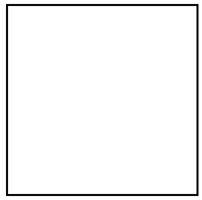

Figure 1.7.1 Humean and Moivrean versions of the event that Dennis goes to Sigmund's house.

The local nature of Humean events, and their fewer alternatives, make them more suitable, in general, for the representation of causes. When Dennis ends up practicing his saxophone at Sigmund's house, and we say that he remembered to practice because he went to Sigmund's house, we are thinking of his going to Sigmund's house as a Humean event, localized to the situation where the other alternatives are going to Alex's house or staying at his own house with Alex. It is not so reasonable to make the causal statement if we regard going to Sigmund's house as a Moivrean event, for then we must weigh also as an alternative the possibility that Alex does not come over, and under this alternative Dennis also has a good chance of practicing.

It may further clarify the difference between Moivrean and Humean events to note that a Moivrean event necessarily either happens or fails, since every path through the tree is either in it or outside it, while a Humean event is more contingent. If it corresponds to a single outcome of a particular local experiment, which may or may not be performed, a Humean event may not even have the opportunity to happen. We can sometimes say that two experiments performed at different places in a tree are the same, but often there are some paths through the tree where the opportunity to perform a given experiment simply does not arise. In Figure 1.1.2, for example, there are two places where the third round is played, but there are some paths on which no third round is played. Similarly, if Alex comes to Dennis's house, Dennis never faces the choice between reading and television. In the immense fragment of nature's tree that we might evoke in order to discuss whether smoking causes heart disease, many different paths would include steps corresponding to Dennis's smoking, but there might be other paths where Dennis does not even exist, perhaps because of his father's early death from lung cancer or heart disease.

The differences between Moivrean and Humean events are not absolute. A Moivrean event is global only with respect to the starting point of the fragment of nature's tree we are considering, and a Humean event can be a single step only with respect to the limited degree of detail provided by the fragment. We can always make the fragment a branch in a larger fragment with an earlier starting point, so that the Moivrean event becomes contingent and relatively more localized. And we can always refine the fragment, so that what was a single step becomes a collection of steps or chains along different paths. In Figure 1.7.2, for example, the step where Nell decides to jog becomes two separate steps when the tree is refined.

Figure 1.7.2 More and less detailed trees. The Humean event that Nell will jog in the morning is a single step only when considered in the less detailed tree.

1.8. Three Levels of Causal Language

There are three important and distinct levels of discourse at which we can use causal language. First, we can use causal language in referring to a particular path through nature's tree, as when we pick out the most salient steps along that path. We say, for example, that Dennis forgot to practice his saxophone because he went to Alex's house. This is the level of singular causation, which is central to the historical sciences. Second, we can use causal language in referring to nature's tree as a whole, as when we say two Moivrean events are independent. This is the level with which this book is primarily concerned. Third, we can use causal language at a yet more general level, to discuss how nature's tree might be manipulated by human intervention. We might consider, for example, whether we can make Dennis remember to practice his saxophone by forbidding him to leave the house or watch television. This is the level of causal law, which is central to the manipulative sciences, engineering and artificial intelligence.

Sometimes it is useful to fold the third level of causal language back into the second level, by shifting attention to a broader version of nature's tree, which includes steps representing the decisions that we take to construct or manipulate the narrower version. If nature has no probabilities for these decisions, then this larger tree may be a decision tree—a tree in which branching probabilities are supplied for only some of the experiments. Figure 1.8.1 shows an example. As we will see in Chapter 12, the theory of causality in nature's probability tree generalizes to nature's decision tree, and concepts within this generalization can take over much of the work sometimes done by the ideas of intervention and causal law.

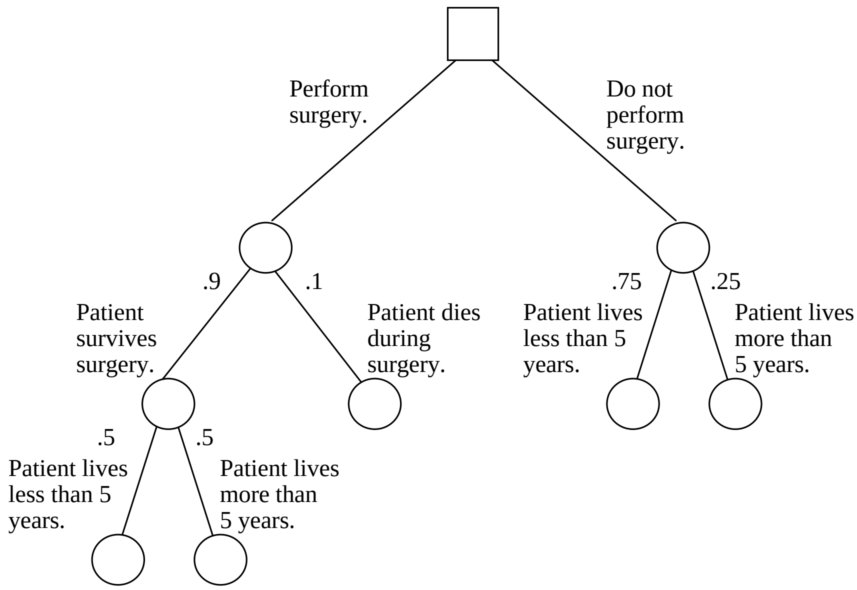

Figure 1.8.1 A decision tree. The square node represents a decision, which is not probabilized.

1.9. An Outline of the Book

Although the preceding discussion has touched on nearly all the remaining chapters of the book, it may still be useful to outline linearly the course of the exposition.

Chapters 2, 3, and 4 provide a mathematical and philosophical foundation for probability trees. Chapters 2 lays out terminology for events and situations, and Chapter 3 formulates the basic ideas of mathematical probability using this terminology. Chapter 4 is more philosophical; it makes the case that probability trees provide a complete framework for probability, capturing the frequency aspects without any further structure of repetition and capturing the subjective aspects without any further structure for changes in belief.

Chapters 5, 6, and 7 study the concepts of independence, tracking, and sign for events, and Chapters 8, 9, and 10 generalize these concepts to variables. Dealing first with events and then with variables entails some repetition; a mathematically more succinct approach would first give the most general definitions and then specialize to the simpler cases. But since the concepts are new, it seems wise to go more slowly, dealing first with the simplest cases. This approach has the advantage, moreover, that it brings out relations between ideas that have been studied separately by different scholarly communities. The concepts that we define for events are closely related to ideas that have been studied most by philosophers, in the tradition initiated by Hans Reichenbach, whereas their generalizations to variables bear more resemblance to ideas that have been studied more often by statisticians.

In Chapter 11, the book moves to a more abstract mathematical level, which is needed for a full appreciation of the generality of probability trees. Chapter 11 itself reformulates the idea of a probability tree in a way that accommodates infinities and also permits deterministic steps, steps where there is no choice how to proceed.

Chapter 12 explores the generalization from probability tree to martingale trees. Martingale trees include probability trees, decision trees, and yet more general trees;

mathematically, their probability structure is specified by a linear space of martingales, which may determine probabilities and expected values for only some events and variables.

Chapter 13 studies the relationship between more and less detailed probability and martingale trees. We distinguish two cases: the case where a tree for an observer is refined to a more detailed tree for the same observer, as in Figure 1.7.2, and the case where a tree for one observer is grounded in a more detailed tree for a more knowledgeable observer.

In Chapter 14, we arrive finally at the topic of causal explanation. Using the ideas developed in Chapters 2 through 13, Chapter 14 develops the theme that causal explanations are partial descriptions of nature's tree. Causal conjecture is the formulation of such descriptions on the basis of limited evidence, and causal inference is the confirmation and testing of causal conjecture.

Chapter 15 uses the work of Chapters 8, 9, and 10 to interpret statistical models, including models consisting of single equations, models consisting of systems of equations, and models displayed as prediction diagrams.

Finally, Chapter 16 explores the computational representation of probability and martingale trees. If causal conjecture is to be a practical guide to action, we must have better methods for representing our conjectures than the pictures of trees used for the examples in this book. Consideration of the possibilities takes us away from conventional mathematics into the domain of intuitionistic logic and give us a broader perspective on the idea of causal law.

Chapters 14 and 15, which are concerned with causal explanation and causal modeling, constitute the heart of the book. They can be read at many different levels, and although a full understanding of them may require a mastery of the ideas in Chapters 2 through 13, most readers should turn to them for a first reading immediately after completing this introduction.