3.1 Introduction

The problems addressed in this chapter can best be illustrated through a classical example due to Cochran (see Wainer 1989). Consider an experiment in which soil fumigants (X) are used to increase oat crop yields (Y) by controlling the eelworm population (Z); the fumigants may also have direct effects (both beneficial and adverse) on yields beside the control of eelworms. We wish to assess the total effect of the fumigants on yields when this typical study is complicated by several factors. First, controlled randomized experiments are unfeasible—farmers insist on deciding for themselves which plots are to be fumigated. Second, farmers' choice of treatment depends on last year's eelworm population (Z_0), an unknown quantity that is strongly correlated with this year's population. Thus we have a classical case of confounding bias that interferes with the assessment of treatment effects regardless of sample size. Fortunately, through laboratory analysis of soil samples, we can determine the eelworm populations before and after the treatment; furthermore, because the fumigants are known to be active for a short period only, we can safely assume that they do not affect the growth of eelworms surviving the treatment. Instead, eelworms' growth depends on the population of birds (and other predators), which is correlated with last year's eelworm population and hence with the treatment itself.

The method developed in this chapter permits the investigator to translate complex considerations of this sort into a formal language and thereby facilitate the following tasks:

1. explicating the assumptions that underlie the model;
2. deciding whether the assumptions are sufficient to obtain consistent estimates of the target quantity—the total effect of the fumigants on yields;
3. providing (if the answer to item 2 is affirmative) a closed-form expression for the target quantity in terms of distributions of observed quantities; and
4. suggesting (if the answer to item 2 is negative) a set of observations and experiments that, if performed, would render a consistent estimate feasible.

The first step in this analysis is to construct a causal diagram like the one given in Figure 3.1, which represents the investigator's under-

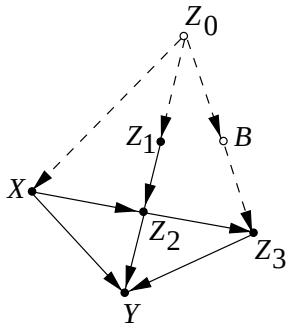

Figure 3.1: A causal diagram representing the effect of fumigants (X) on yields (Y).

standing of the major causal influences among measurable quantities in the domain. For example, the quantities Z_1, Z_2, Z_3 represent the eelworm population before treatment, after treatment, and at the end of the season, respectively. The Z_0 term represents last year's eelworm population; because it is an unknown quantity, it is denoted by a hollow circle, as is the quantity B, the population of birds and other predators. Links in the diagram are of two kinds: those that connect unmeasured quantities are designated by dashed arrows, those connecting measured quantities by solid arrows. The substantive assumptions embodied in the diagram are negative causal assertions which are conveyed through the links *missing* from the diagram. For example, the missing arrow between Z_1 and Y signifies the investigator's understanding that pretreatment eelworms can not affect oat plants directly; their entire influence on oat yields is mediated by the posttreatment conditions, Z_2 and Z_3. Our purpose is not to validate or repudiate such domain-specific assumptions but rather to test whether a given set of assumptions is sufficient for quantifying causal effects from nonexperimental data—here, estimating the total effect of fumigants on yields.

The causal diagram in Figure 3.1 is similar in many respects to the path diagrams devised by Wright (1921); both reflect the investigator's subjective and qualitative knowledge of causal influences in the domain, both employ directed acyclic graphs, and both allow for the incorpora-

tion of latent or unmeasured quantities. The major differences lie in the method of analysis. First, whereas path diagrams have been analyzed mostly in the context of linear models with Gaussian noise, causal diagrams permit arbitrary nonlinear interactions. In fact, our analysis of causal effects will be entirely nonparametric, entailing no commitment to a particular functional form for equations and distributions. Second, causal diagrams will be used not only as a passive language to communicate assumptions but also as an active computational device through which the desired quantities are derived. For example, the method to be described allows an investigator to inspect the diagram of Figure 3.1 and make the following immediate conclusions.

1. The total effect of X on Y can be estimated consistently from the observed distribution of $X,\ Z_1,\ Z_2,\ Z_3$, and Y.

2. The total effect of X on Y (assuming discrete variables throughout) is given by the formula[1]

$$P(y|\hat{x}) = \sum_{z_1}\sum_{z_2}\sum_{z_3} P(y|z_2,\ z_3,\ x)P(z_2|z_1,\ x) \times \sum_{x'} P(z_3|z_1,\ z_2,\ x')P(z_1,\ x'), \tag{3.1}$$

   where $P(y|\hat{x})$ stands for the probability of achieving a yield level of $Y = y$, given that the treatment is set to level $X = x$ by external intervention.

3. A consistent estimation of the total effect of X on Y would not be feasible if Y were confounded with Z_3; however, confounding Z_2 and Y will not invalidate the formula for $P(y|\hat{x})$.

These conclusions will be obtained either by analyzing the graphical properties of the diagram or by performing a sequence of symbolic derivations (governed by the diagram) that gives rise to causal effect formulas such as (3.1).

[1] The notation $P_x(y)$ was used in Chapter 1; it is changed henceforth to $P(y|\hat{x})$ or $P(y|do(x))$ because of the inconvenience in handling subscripts. The reader need not be intimidated if, at this point, (3.1) appears unfamiliar. After reading Section 3.4, the reader should be able to derive such formulas with greater ease than solving algebraic equations. Note that x' is merely an index of summation that ranges over the values of X.