Page 2

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.

Page 3

- 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.
- The total effect of X on Y (assuming discrete variables throughout) is given by the formula

$$\begin{aligned} P(y \mid \hat{x}) = & \sum_{z_1} \sum_{z_2} \sum_{z_3} P\left(y \mid z_2, z_3, x\right) P\left(z_2 \mid z_1, x\right) \\ & \times \sum_{x'} P\left(z_3 \mid z_1, z_2, x'\right) P\left(z_1, x'\right) \end{aligned}$$

- 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 $P(y \mid \hat{x})$

Page 4

If any of these factors is judged to be influencing two or more variables (thus violating the independence assumption), then that factor must enter the analysis as an unmeasured (or latent) variable and be represented in the graph by a hollow node, such as Z_0 and B in Figure 3.1.

$$\begin{array}{ll} Z_0 = f_0\left(\varepsilon_0\right), & B = f_B\left(Z_0, \varepsilon_B\right) \\ Z_1 = f_1\left(Z_0, \varepsilon_1\right), & X = f_X\left(Z_0, \varepsilon_X\right) \\ Z_2 = f_2\left(X, Z_1, \varepsilon_2\right), & Y = f_Y\left(X, Z_2, Z_3, \varepsilon_Y\right) \\ Z_3 = f_3\left(B, Z_2, \varepsilon_3\right) & \end{array}$$

Page 5

Our example of Figure 3.1, yields

$$\begin{aligned} P\left(z_0, x, z_1, b, z_2, z_3, y\right) = & P\left(z_0\right) P\left(x \mid z_0\right) P\left(z_1 \mid z_0\right) P\left(b \mid z_0\right) \\ & \times P\left(z_2 \mid x, z_1\right) P\left(z_3 \mid z_2, b\right) P\left(y \mid x, z_2, z_3\right). \end{aligned}$$

The model described in Figure 3.1 is semi-Markovian if the observed variables are $\{X, Y, Z_1, Z_2, Z_3\}$; it would turn Markovian if Z_0 and B were observed as well.

The equational model is the nonparametric analog of the so-called structural equation model (Wright 1921; Goldberger 1973), except that the functional form of the equations (as well as the distribution of the disturbance terms) will remain unspecified.

The equality signs in structural equations convey the asymmetrical counterfactual relation of "is determined by," and each equation represents a stable autonomous mechanism.

Effect of Interventions

For example, the intervention $do\,(X = x')$ will transform the pre-intervention distribution given into the product.

$$\begin{aligned} P\left(z_0, z_1, b, z_2, z_3, y \mid \hat{x}'\right) =& P\left(z_0\right) P\left(z_1 \mid z_0\right) P\left(b \mid z_0\right) \\ & \times P\left(z_2 \mid x', z_1\right) P\left(z_3 \mid z_2, b\right) P\left(y \mid x', z_2, z_3\right). \end{aligned}$$

Backdoor Criterion

In Figure 3.4, for example, the sets $Z_1 = \{X_3, X_4\}$ and $Z_2 = \{X_4, X_5\}$ meet the back-door criterion, but $Z_3 = \{X_4\}$ does not because X_4 does not block the path $(X_i, X_3, X_1, X_4, X_2, X_5, X_j)$.