Chapter 8

Imperfect Experiments: Bounding Effects and Counterfactuals

> *Would that I could discover truth*
> *as easily as I can uncover falsehood.*
> Cicero (44 B.C.)

Preface

In this chapter we describe how graphical and counterfactual models (Sections 3.2 and 7.1) can combine to elicit causal information from imperfect experiments: experiments that deviate from the ideal protocol of randomized control. A common deviation occurs, for example, when subjects in a randomized clinical trial do not fully comply with their assigned treatment, thus compromising the identification of causal effects. When conditions for identification are not met, the best one can do is derive *bounds* for the quantities of interest—namely, a range of possible values that represents our ignorance about the data-generating process and that cannot be improved with increasing sample size. The aim of this chapter is to demonstrate (i) that such bounds can be derived by simple algebraic methods and (ii) that, despite the imperfection of the experiments, the derived bounds can yield significant and sometimes accurate information on the impact of a policy on the entire population as well as on a particular individual who participated in the study.

8.1 Introduction

8.1.1 Imperfect and Indirect Experiments

Standard experimental studies in the biological, medical, and behavioral sciences invariably invoke the instrument of randomized control; that is, subjects are assigned at random to various groups (or treatments or programs), and the mean differences between participants in different groups are regarded as measures of the efficacies of the associated programs. Deviations from this ideal setup may take place either by failure to meet any of the experimental requirements or by deliberate attempts to relax these requirements. *Indirect experiments* are studies in which randomized control is either unfeasible or undesirable. In such experiments, subjects are still assigned at random to various groups, but members of each group are simply encouraged (rather than forced) to participate in the program associated with the group; it is up to the individuals to select among the programs.

Recently, use of strict randomization in social and medical experimentation has been questioned for three major reasons.

1. Perfect control is hard to achieve or ascertain. Studies in which treatment is assumed to be randomized may be marred by uncontrolled *imperfect compliance.* For example, subjects experiencing adverse reactions to an experimental drug may decide to reduce the assigned dosage. Alternatively, if the experiment is testing a drug for a terminal disease, a subject suspecting that he or she is in the control group may obtain the drug from other sources. Such imperfect compliance renders the experiment indirect and introduces bias into the conclusions that researchers draw from the data. This bias cannot be corrected unless detailed models of compliance are constructed (Efron and Feldman 1991).

2. Denying subjects assigned to certain control groups the benefits of the best available treatment has moral and legal ramifications. For example, in AIDS research it is difficult to justify placebo programs because those patients assigned to the placebo group would be denied access to potentially life-saving treatment (Palca 1989).

3. Randomization, by its very presence, may influence participation as well as behavior (Heckman 1992). For example, eligible candidates may be wary of applying to a school once they discover that it deliberately randomizes its admission criteria. Likewise, as Kramer and Shapiro (1984) noted, subjects in drug trials may be less likely to participate in randomized trials than in nonexperimental studies, even when the treatments are equally nonthreatening.

Altogether, researchers are beginning to acknowledge that mandated randomization may undermine the reliability of experimental evidence and that experimentation with human subjects often involves—and sometimes *should* involve—an element of self-selection.

This chapter concerns the drawing of inferences from studies in which subjects have final choice of program; the randomization is confined to an indirect *instrument* (or *assignment*) that merely encourages or discourages participation in the various programs. For example, in evaluating the efficacy of a given training program, notices of eligibility may be sent to a randomly selected group of students or, alternatively, eligible candidates may be selected at random to receive scholarships for participating in the program. Similarly, in drug trials, subjects may be given randomly chosen advice on recommended dosage level, yet the final choice of dosage will be determined by the subjects to fit their individual needs.

Imperfect compliance poses a problem because simply comparing the fractions in the treatment and control groups may provide a misleading estimate for how effective the treatment would be if applied uniformly to the population. For example, if those subjects who declined to take the drug are precisely those who would have responded adversely, the experiment might conclude that the drug is more effective than it actually is. In Chapter 3 (see Section 3.5, Figure 3.7(b)), we showed that treatment effectiveness in such studies is actually *non-identifiable*. That is, in the absence of additional modeling assumptions, treatment effectiveness cannot be estimated from the data without bias, even when the number of subjects in the experiment approaches infinity and even when a record is available of the action and response of each subject.

The question we attempt to answer in this chapter is whether indirect randomization can provide information that allows approximate assessment of the intrinsic merit of a program, as would be measured, for example, if the program were to be extended and mandated uniformly to the population. The analysis presented shows that, given a minimal set of assumptions, such inferences are indeed possible—albeit in the form of *bounds*, rather than precise point estimates, for the causal effect of the program or treatment. These bounds can be used by the analyst to guarantee that the causal effect of a given program must be higher than one measurable quantity and lower than another.

Our most crucial assumption is that, for any given person, the encouraging instrument influences the treatment chosen by that person but has no effect on how that person would respond to the treatment chosen (see the definition of instrumental variables in Section 7.4.5). The second assumption, one which is always made in experimental studies, is that subjects respond to treatment independently of one other. Other than these two assumptions, our model places no constraints on how tendencies to respond to treatments may interact with choices among treatments.

8.1.2 Noncompliance and Intent to Treat

In a popular compromising approach to the problem of imperfect compliance, researchers perform an "intent to treat" analysis in which the control and treatment group are compared without regard to whether the treatment was actually received.[1] The result of such an analysis is a measure of how well the treatment *assignment* affects the disease, as opposed to the desired measure of how well the treatment *itself* affects the disease. Estimates based on intent-to-treat analyses are valid only as long as the experimental conditions perfectly mimic the conditions prevailing in the eventual usage of the treatment. In particular, the experiment should mimic subjects' incentives for receiving each treatment. In situations where field incentives are more compelling than experimental incentives, as is usually the case when drugs receive the approval of a government agency, treatment effectiveness may vary sig-

[1] This approach is currently used by the FDA to approve new drugs.

nificantly from assignment effectiveness. For example, imagine a study in which (a) the drug has an adverse effect on a large segment of the population and (b) only those members of the segment who drop from the treatment "arm" (subpopulation) recover. The intent-to-treat analysis will attribute these cases of recovery to the drug because they are part of the intent-to-treat arm, although in reality these cases recovered by *avoiding* the treatment.

Another approach to the problem is to use a correction factor based on an instrumental variables formula (Angrist et al. 1996), according to which the intent-to-treat measure should be divided by the fraction of subjects who comply with the treatment assigned to them. Angrist et al. (1996) showed that, under certain conditions, the corrected formula is valid for the subpopulation of "responsive" subjects—that is, subjects who would have changed treatment status if given a different assignment. Unfortunately, this subpopulation cannot be identified and, more seriously, it cannot serve as a basis for policies involving the entire population because it is instrument-dependent: individuals who are responsive in the study may not remain responsive in the field, where the incentives for obtaining treatment differ from those used in the study. We therefore focus our analysis on the stable aspect of the treatment—the aspect that would remain invariant to changes in compliance behavior.

8.2 Bounding Causal Effects

8.2.1 Problem Formulation

The basic experimental setting associated with indirect experimentation is shown in Figure 8.1, which is isomorphic to Figures 3.7(b) and 5.9. To focus the discussion, we will consider a prototypical clinical trial with partial compliance, although in general the model applies to any study in which a randomized instrument encourages subjects to choose one program over another.

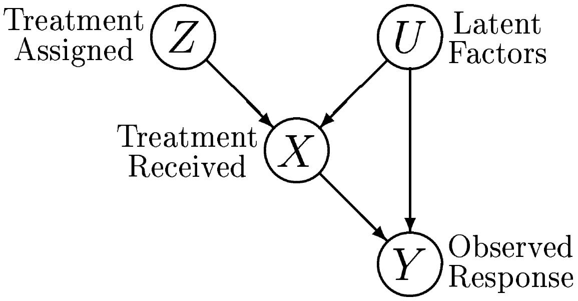


Figure 8.1: Graphical representation of causal dependencies in a randomized clinical trial with partial compliance.

We assume that Z, X, Y are observed binary variables, where Z represents the (randomized) treatment assignment, X is the treatment actually received, and Y is the observed response. The U term represents all factors, both observed and unobserved, that influence the way a subject responds to treatments; hence, an arrow is drawn from U to Y. The arrow from U to X denotes that the U factors may also influence the subject's choice of treatment X; this dependence may represent a complex decision process standing between the assignment (Z) and the actual treatment (X).

To facilitate the notation, we let z, x, y represent (respectively) the values taken by the variables Z, X, Y, with the following interpretation:

$z \in \{z_0, z_1\}$, z_1 asserts that treatment has been assigned (z_0, its negation);

$x \in \{x_0, x_1\}$, x_1 asserts that treatment has been administered (x_0, its negation); and

$y \in \{y_0, y_1\}$, y_1 asserts a positive observed response (y_0, its negation).

The domain of U remains unspecified and may, in general, combine the spaces of several random variables, both discrete and continuous.

The graphical model reflects two assumptions.

1. The assigned treatment Z does not influence Y directly but rather through the actual treatment X. In practice, any direct effect Z might have on Y would be adjusted for through the use of a placebo.

2. The variables Z and U are marginally independent; this is ensured through the randomization of Z, which rules out a common cause for both Z and U.

These assumptions impose on the joint distribution the decomposition

$$P(y, x, z, u) \;=\; P(y|x, u)P(x|z, u)P(z)P(u), \qquad (8.1)$$

which, of course, cannot be observed directly because U is unobserved. However, the marginal distribution $P(y, x, z)$ and, in particular, the conditional distributions

$$P(y, x|z) = \sum_u P(y|x, u)P(x|z, u)P(u), \quad z \in \{z_0, z_1\}, \qquad (8.2)$$

are observed,[2] and the challenge is to assess from these distributions the average *change* in Y due to treatment.

Treatment effects are governed by the distribution $P(y|do(x))$, which—using the truncated factorization formula of (3.10)—is given by

$$P(y|do(x)) = \sum_u P(y|x, u)P(u); \qquad (8.3)$$

[2] In practice, of course, only a finite sample of $P(y, x|z)$ will be observed. But our task is one of identification, not estimation, so we make the large-sample assumption and consider $P(y, x|z)$ as given.

here, the factors $P(y|x,u)$ and $P(u)$ are the same as those in (8.2). Therefore, if we are interested in the average change in Y due to treatment then we should compute the *average causal effect*, ACE$(X \rightarrow Y)$ (Holland 1988), which is given by

$$\begin{aligned} \text{ACE}(X \rightarrow Y) &= P(y_1|do(x_1)) - P(y_1|do(x_0)) \\ &= \sum_u [P(y_1|x_1,u) - P(y_1|x_0,u)P(u)]. \end{aligned} \tag{8.4}$$

Our task is then to estimate or bound the expression in (8.4) given the observed probabilities $P(y,x|z_0)$ and $P(y,x|z_1)$, as expressed in (8.2). This task amounts to a constrained optimization exercise of finding the highest and lowest values of (8.4) subject to the equality constraint in (8.2), where the maximization ranges over all possible functions

$$P(u),\ P(y_1|x_0,u),\ P(y_1|x_1,u),\ P(x_1|z_0,u),\ \text{and}\ P(x_1|z_1,u)$$

that satisfy those constraints.

8.2.2 The Evolution of Potential-Response Variables

The bounding exercise described in Section 8.2.1 can be solved using conventional techniques of mathematical optimization. However, the continuous nature of the functions involved—as well as the unspecified domain of U—makes this representation inconvenient for computation. Instead, we can use the observation that U can always be replaced by a finite-state variable such that the resulting model is equivalent with respect to all observations and manipulations of Z, X, Y (Pearl 1994a).

Consider the structural equation that connects two binary variables, Y and X, in a causal model:

$$y = f(x,u).$$

For any given u, the relationship between X and Y must be one of four functions:

$$\begin{aligned} &f_0 : y = 0, && f_1 : y = x, \\ &f_2 : y \neq x, && f_3 : y = 1. \end{aligned} \tag{8.5}$$

As u varies along its domain, regardless of how complex the variation, the only effect it can have on the model is to switch the relationship between X and Y among these four functions. This partitions the domain of U into four equivalence classes, as shown in Figure 8.2, where each class contains those points u that correspond to the same function. We can thus replace U by a four-state variable, $R(u)$, such that each state represents one of the four functions. The probability $P(u)$ would automatically translate into a probability function $P(r),\ r = 0, 1, 2, 3$, that is given by the total weight assiged to the equivalence class corresponding to r. A state-minimal variable like R is called a "response variable" by Balke and Pearl (1994a, b) and a "mapping variable" by Heckerman and Shachter (1995).[3]

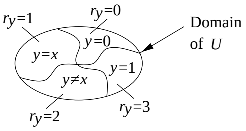


Figure 8.2: The partition of U into four equivalence classes, each inducing a distinct functional mapping from X to Y for any given function $y = f(x, u)$.

Because Z, X, and Y are all binary variables, the state space of U divides into 16 *equivalence classes*: each class dictates two functional mappings, one from Z to X and the other from X to Y. To describe these equivalence classes, it is convenient to regard each of them as a

[3] In the potential-outcome model (see Section 7.4.4), u stands for an experimental unit and $R(u)$ corresponds to the potential response of unit u to treatment x. The assumption that each experimental unit (e.g., an individual subject) possesses an intrinsic, seemingly "fatalistic" response function has met with some objections (Dawid 1997), owing to the complexity and inherent unobservability of the many factors that might govern an individual response to treatment. The equivalence-class formulation of $R(u)$ mitigates those objections by showing that $R(u)$ evolves naturally and mathematically from any complex system of stochastic latent variables, provided only that we acknowledge the existence of such variables through the equation $y = f(x, u)$. Those who invoke quantum-mechanical objections to the latter step as well (e.g. Salmon 1998), should regard the functional relationship $y = f(x, u)$ as an abstract mathematical construct, representing the extreme points (vertices) of the set of conditional probabilities $P(y|x, u)$ satisfying the constraints of (8.1) and (8.2).

point in the joint space of two four-valued variables R_x and R_y. The variable R_x determines the compliance behavior of a subject through the mapping

$$x = f_X(z, r_x) = \begin{cases} x_0 & \text{if} \quad r_x = 0; \\ x_0 & \text{if} \quad r_x = 1 \quad \text{and} \quad z = z_0, \\ x_1 & \text{if} \quad r_x = 1 \quad \text{and} \quad z = z_1; \\ x_1 & \text{if} \quad r_x = 2 \quad \text{and} \quad z = z_0, \\ x_0 & \text{if} \quad r_x = 2 \quad \text{and} \quad z = z_1; \\ x_1 & \text{if} \quad r_x = 3. \end{cases} \tag{8.6}$$

Imbens and Rubin (1997) call a subject with compliance behavior $r_x = 0, 1, 2, 3$ (respectively) a *never-taker*, a *complier*, a *defier*, and an *always-taker*. Similarly, the variable R_y determines the response behavior of a subject through the mapping:

$$y = f_Y(x, r_y) = \begin{cases} y_0 & \text{if} \quad r_y = 0; \\ y_0 & \text{if} \quad r_y = 1 \quad \text{and} \quad x = x_0, \\ y_1 & \text{if} \quad r_y = 1 \quad \text{and} \quad x = x_1; \\ y_1 & \text{if} \quad r_y = 2 \quad \text{and} \quad x = x_0, \\ y_0 & \text{if} \quad r_y = 2 \quad \text{and} \quad x = x_1; \\ y_1 & \text{if} \quad r_y = 3. \end{cases} \tag{8.7}$$

Following Heckerman and Shachter (1995), we call the response behavior $r_y = 0, 1, 2, 3$ (respectively) *never-recover*, *helped*, *hurt*, and *always-recover*.

The correspondence between the states of variable R_y and the potential response variables, Y_{x_0} and Y_{x_1}, defined in Section 7.1 (Definition 7.1.4) is as follows:

$$Y_{x_1} = \begin{cases} y_1 & \text{if } r_y = 1 \text{ or } r_y = 3, \\ y_0 & \text{otherwise;} \end{cases}$$

$$Y_{x_0} = \begin{cases} y_1 & \text{if } r_y = 2 \text{ or } r_y = 3, \\ y_0 & \text{otherwise.} \end{cases}$$

In general, response and compliance may not be independent, hence the double arrow $R_x \dashleftarrow\dashrightarrow R_y$ in Figure 8.3. The joint distribution over $R_x \times R_y$ requires 15 independent parameters, and these parameters

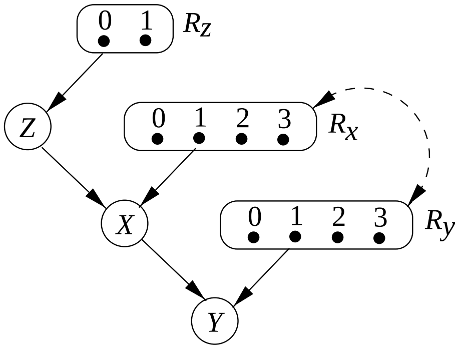


Figure 8.3: A structure equivalent to that of Figure 8.1 but employing finite-state response variables R_z, R_x, and R_y.

are sufficient for specifying the model of Figure 8.3, $P(y, x, z, r_x, r_y) = P(y|x, r_y)P(x|r_x, z)P(z)P(r_x, r_y)$, because Y and X stand in fixed functional relations to their parents in the graph. The causal effect of the treatment can now be obtained directly from (8.7), giving

$$P(y_1|do(x_1)) = P(r_y{=}1) + P(r_y{=}3), \tag{8.8}$$
$$P(y_1|do(x_0)) = P(r_y{=}2) + P(r_y{=}3), \tag{8.9}$$

and

$$\text{ACE}(X \to Y) = P(r_y{=}1) - P(r_y{=}2). \tag{8.10}$$

8.2.3 Linear Programming Formulation

By explicating the relationship between the parameters of $P(y, x|z)$ and those of $P(r_x, r_y)$, we obtain a set of linear constraints needed for minimizing or maximizing ACE$(X \to Y)$ given $P(y, x|z)$.

The conditional distribution $P(y, x|z)$ over the observable variables is fully specified by eight parameters, which will be written as follows:

$$\begin{array}{ll} p_{00.0} = P(y_0, x_0|z_0), & p_{00.1} = P(y_0, x_0|z_1), \\ p_{01.0} = P(y_0, x_1|z_0), & p_{01.1} = P(y_0, x_1|z_1), \\ p_{10.0} = P(y_1, x_0|z_0), & p_{10.1} = P(y_1, x_0|z_1), \\ p_{11.0} = P(y_1, x_1|z_0), & p_{11.1} = P(y_1, x_1|z_1). \end{array}$$

The probabilistic constraints

$$\sum_{n=00}^{11} p_{n.0} = 1 \quad \text{and} \quad \sum_{n=00}^{11} p_{n.1} = 1 \tag{8.11}$$

further imply that $\vec{p} = (p_{00.0}, \ldots, p_{11.1})$ can be specified by a point in 6-dimensional space. This space will be referred to as P.

The joint probability $P(r_x, r_y)$ has 16 parameters:

$$q_{jk} \;=\; P(r_x{=}j,\ r_y{=}k),$$

where $j, k \in \{0, 1, 2, 3\}$. The probabilistic constraint

$$\sum_{j=0}^{3}\sum_{k=0}^{3} q_{jk} \;=\; 1$$

implies that $\vec{q}$ specifies a point in 15-dimensional space. This space will be referred to as Q.

Equation (8.10) can now be rewritten as a linear combination of the Q parameters:

$$\text{ACE}(X \to Y) = q_{01} + q_{11} + q_{21} + q_{31} - q_{02} - q_{12} - q_{22} - q_{32}. \tag{8.12}$$

Applying (8.6) and (8.7) we can write the linear transformation from a point $\vec{q}$ in Q to a point $\vec{p}$ in P:

$$\begin{aligned} p_{00.0} &= q_{00} + q_{01} + q_{10} + q_{11}, & p_{00.1} &= q_{00} + q_{01} + q_{20} + q_{21},\\ p_{01.0} &= q_{20} + q_{22} + q_{30} + q_{32}, & p_{01.1} &= q_{10} + q_{12} + q_{30} + q_{32},\\ p_{10.0} &= q_{02} + q_{03} + q_{12} + q_{13}, & p_{10.1} &= q_{02} + q_{03} + q_{22} + q_{23},\\ p_{11.0} &= q_{21} + q_{23} + q_{31} + q_{33}, & p_{11.1} &= q_{11} + q_{13} + q_{31} + q_{33}; \end{aligned}$$

this can be written in matrix form as $\vec{p} = \boldsymbol{R}\vec{q}$.

Given a point $\vec{p}$ in P-space, the strict lower bound on $\text{ACE}(X \to Y)$ can be determined by solving the following linear programming problem.

Minimize $q_{01} + q_{11} + q_{21} + q_{31} - q_{02} - q_{12} - q_{22} - q_{32}$
subject to:

$$\begin{aligned} \sum_{j=0}^{3}\sum_{k=0}^{3} q_{jk} &= 1,\\ \boldsymbol{R}\vec{q} &= \vec{p},\\ q_{jk} &\geq 0 \text{ for } j, k \in \{0, 1, 2, 3\}. \end{aligned} \tag{8.13}$$

Moreover, for problems of this size, procedures are available for deriving symbolic expressions as well (Balke 1995), leading to the following lower bound on the treatment effect

$$\text{ACE}(X \to Y) \geq \max \left\{ \begin{array}{c} p_{11.1} + p_{00.0} - 1 \\ p_{11.0} + p_{00.1} - 1 \\ p_{11.0} - p_{11.1} - p_{10.1} - p_{01.0} - p_{10.0} \\ p_{11.1} - p_{11.0} - p_{10.0} - p_{01.1} - p_{10.1} \\ -p_{01.1} - p_{10.1} \\ -p_{01.0} - p_{10.0} \\ p_{00.1} - p_{01.1} - p_{10.1} - p_{01.0} - p_{00.0} \\ p_{00.0} - p_{01.0} - p_{10.0} - p_{01.1} - p_{00.1} \end{array} \right\} .\text{(8.14a)}$$

Similarly, the upper bound is given by

$$\text{ACE}(X \to Y) \leq \min \left\{ \begin{array}{c} 1 - p_{01.1} - p_{10.0} \\ 1 - p_{01.0} - p_{10.1} \\ -p_{01.0} + p_{01.1} + p_{00.1} + p_{11.0} + p_{00.0} \\ -p_{01.1} + p_{11.1} + p_{00.1} + p_{01.0} + p_{00.0} \\ p_{11.1} + p_{00.1} \\ p_{11.0} + p_{00.0} \\ -p_{10.1} + p_{11.1} + p_{00.1} + p_{11.0} + p_{10.0} \\ -p_{10.0} + p_{11.0} + p_{00.0} + p_{11.1} + p_{10.1} \end{array} \right\} \text{(8.14b)}$$

We may also derive bounds for (8.8) and (8.9) individually (under the same linear constraints), giving:

$$P(y_1|do(x_0)) \geq \max \left\{ \begin{array}{c} p_{10.0} + p_{11.0} - p_{00.1} - p_{11.1} \\ p_{10.1} \\ p_{10.0} \\ p_{01.0} + p_{10.0} - p_{00.1} - p_{01.1} \end{array} \right\},$$

$$P(y_1|do(x_0)) \leq \min \left\{ \begin{array}{c} p_{01.0} + p_{10.0} + p_{10.1} + p_{11.1} \\ 1 - p_{00.1} \\ 1 - p_{00.0} \\ p_{10.0} + p_{11.0} + p_{01.1} + p_{10.1} \end{array} \right\}; \tag{8.15}$$

$$P(y_1|do(x_1)) \geq \max \left\{ \begin{array}{c} p_{11.0} \\ p_{11.1} \\ -p_{00.0} - p_{01.0} + p_{00.1} + p_{11.1} \\ -p_{01.0} - p_{10.0} + p_{10.1} + p_{11.1} \end{array} \right\},$$

$$P(y_1|do(x_1)) \leq \min \left\{ \begin{array}{c} 1 - p_{01.1} \\ 1 - p_{01.0} \\ p_{00.0} + p_{11.0} + p_{10.1} + p_{11.1} \\ p_{10.0} + p_{11.0} + p_{00.1} + p_{11.1} \end{array} \right\}. \qquad (8.16)$$

These expressions give the tightest possible assumption-free[4] bounds on the quantities involved.

8.2.4 The Natural Bounds

The expression for ACE$(X \rightarrow Y)$ (equation (8.4)) can be bounded by two simple formulas, each made up of the first two terms in (8.14a and (8.14b)) (Robins 1989; Manski 1990; Pearl 1994):

$$\begin{aligned} \text{ACE}(X \rightarrow Y) &\geq P(y_1|z_1) - P(y_1|z_0) - P(y_1, x_0|z_1) - P(y_0, x_1|z_0) \\ \text{ACE}(X \rightarrow Y) &\leq P(y_1|z_1) - P(y_1|z_0) + P(y_0, x_0|z_1) + P(y_1, x_1|z_0) \end{aligned} \qquad (8.17)$$

Because of their simplicity and wide range of applicability, the bounds given by (8.17) were named the *natural* bounds (Balke and Pearl 1997). The natural bounds guarantee that the causal effect of the actual treatment cannot be smaller than that of the encouragement $(P(y_1|z_1) - P(y_1|z_0))$ by more than the sum of two measurable quantities, $P(y_1, x_0|z_1) + P(y_0, x_1|z_0)$; they also guarantee that the causal effect of the treatment cannot exceed that of the encouragement by more than the sum of two other measurable quantities, $P(y_0, x_0|z_1) + P(y_1, x_1|z_0)$. The width of the natural bounds, not surprisingly, is given by the rate of noncompliance, $P(x_1|z_0) + P(x_0|z_1)$.

The width of the sharp bounds in (8.14ab) can be substantially narrower though. In Balke (1995) and Pearl (1995b), it is shown that—even under conditions of 50% noncompliance—these bounds may collapse to a point and thus permit consistent estimation of ACE$(X \rightarrow Y)$. This occurs whenever (a) the percentage of subjects complying with assignment z_0 is the same as those complying with z_1 and (b) Y and Z

[4] "Assumption-transparent" might be a better term; we make no assumptions about factors that determine subjects' compliance, but we rely on the assumptions of (i) randomized assignment and (ii) no side effects, as displayed in the graph (e.g., Figure 8.1).

are perfectly correlated in at least one treatment arm x (see Table 8.1 in Section 8.5).

Although more complicated than the natural bounds of (8.17ab), the sharp bounds of (8.14ab) are nevertheless easy to assess once we have the frequency data in the eight cells of $P(y, x|z)$. It can also be shown (Balke 1995) that the natural bounds are optimal when we can safely assume that no subject is *contrarian*—iin other words, that no subject would consistently choose a treatment arm contrary to the one assigned.

Note that, if the response Y is continuous, then one can associate y_1 and y_0 with the binary events $Y > t$ and $Y \leq t$ (respectively) and let t vary continuously over the range of Y. (8.15) and (8.16) would then provide bounds on the entire distribution of the treatment effect $P(Y < t \mid do(x))$.

8.2.5 Effect of Treatment on the Treated

Much of the literature assumes that ACE$(X \rightarrow Y)$ is the parameter of interest, because ACE$(X \rightarrow Y)$ predicts the impact of applying the treatment uniformly (or randomly) over the population. However, if a policy maker is not interested in introducing new treatment policies but rather in deciding whether to maintain or terminate an existing program under its current incentive system, then the parameter of interest should measure the impact of the treatment *on the treated*, namely, the mean response of the treated subjects compared to the mean response of these same subjects had they not been treated (Heckman 1992). The appropriate formula for this parameter is

$$\begin{aligned} \text{ACE}^*(X \rightarrow Y) &= P(Y_{x_1} = y_1|x_1) - P(Y_{x_0} = y_1|x_1) \\ &= \sum_u [P(y_1|x_1, u) - P(y_1|x_0, u)]P(u|x_1), \quad (8.18) \end{aligned}$$

which is similar to (8.4) except for replacing the expectation over u with the conditional expectation given $X = x_1$.

The analysis of ACE$^*(X \rightarrow Y)$ reveals that, under conditions of *no intrusion* (i.e., $P(x_1|z_0) = 0$, as in most clinical trials), ACE$^*(X \rightarrow Y)$ can be identified precisely (Bloom 1984; Angrist and Imbens 1991).

The natural bounds governing $\mathrm{ACE}^*(X \to Y)$ in the general case can be obtained by similar means, which yield

$$\begin{aligned}\mathrm{ACE}^*(X \to Y) &\geq \frac{P(y_1|z_1) - P(y_1|z_0)}{P(x_1)/P(z_1)} - \frac{P(y_0, x_1|z_0)}{P(x_1)},\\ \mathrm{ACE}^*(X \to Y) &\leq \frac{P(y_1|z_1) - P(y_1|z_0)}{P(x_1)/P(z_1)} + \frac{P(y_1, x_1|z_0)}{P(x_1)}.\end{aligned} \tag{8.19}$$

The sharp bounds are presented in Balke (1995, p. 113). Clearly, in situations where treatment may be obtained only by those encouraged (by assignment), we have $P(x_1|z_0) = 0$ and

$$\mathrm{ACE}^*(X \to Y) \ = \ \frac{P(y_1|z_1) - P(y_1|z_0)}{P(x_1|z_1)}. \tag{8.20}$$

Unlike $\mathrm{ACE}(X \to Y)$, $\mathrm{ACE}^*(X \to Y)$ is not an intrinsic property of the treatment, since it varies with the encouraging instrument. Hence, its significance lies in studies where it is desired to evaluate the efficacy of an existing program on its current participants.

8.2.6 Example: The Effect of Cholestyramine

To demonstrate by example how the bounds for $\mathrm{ACE}(X \to Y)$ can be used to provide meaningful information about causal effects, consider the Lipid Research Clinics Coronary Primary Prevention Trial data (Program 1984). A portion (covering 337 subjects) of this data was analyzed in Efron and Feldman (1991) and is the focus of this example. Subjects were randomized into two treatment groups of roughly equal size; in one group, all subjects were prescribed cholestyramine (z_1), while subjects in the other group were prescribed a placebo (z_0). Over several years of treatment, each subject's cholesterol level was measured many times, and the average of these measurements was used as the posttreatment cholesterol level (continuous variable C_F). The compliance of each subject was determined by tracking the quantity of prescribed dosage consumed (a continuous quantity).

In order to apply the bounds of (8.17) to data from this study, the continuous data is first transformed, using thresholds, to binary variables representing treatment assignment (Z), received treatment (X), and treatment response (Y). The threshold for dosage consumption was selected as roughly the midpoint between minimum and maximum consumption; the threshold for cholesterol level reduction was set at 28 units. After this "thresholding" procedure, the data samples give rise to the following eight probabilities:[5]

$$\begin{array}{rclrcl}
P(y_0, x_0|z_0) & = & 0.919, & P(y_0, x_0|z_1) & = & 0.315, \\
P(y_0, x_1|z_0) & = & 0.000, & P(y_0, x_1|z_1) & = & 0.139, \\
P(y_1, x_0|z_0) & = & 0.081, & P(y_1, x_0|z_1) & = & 0.073, \\
P(y_1, x_1|z_0) & = & 0.000, & P(y_1, x_1|z_1) & = & 0.473.
\end{array}$$

These data represent a compliance rate of

$$P(x_1|z_1) = 0.139 + 0.473 = 0.61,$$

a mean difference (using $P(z_1) = 0.50$) of

$$P(y_1|x_1) - p(y_1|x_0) = \frac{0.473}{0.473 + 0.139} - \frac{0.073 + 0.081}{1 + 0.315 + 0.073} = 0.662,$$

and an encouragement effect (intent to treat) of

$$P(y_1|z_1) - P(y_1|z_0) = 0.073 + 0.473 - 0.081 = 0.465.$$

According to (8.17), ACE($X \to Y$) can be bounded by

$$\begin{array}{rcl}
\text{ACE}(X \to Y) & \geq & 0.465 - 0.073 - 0.000 = 0.392, \\
\text{ACE}(X \to Y) & \leq & 0.465 + 0.315 + 0.000 = 0.780.
\end{array}$$

These are remarkably informative bounds: although 38.8% of the subjects deviated from their treatment protocol, the experimenter can categorically state that, when applied uniformly to the population, the treatment is guaranteed to increase by at least 39.2% the probability of reducing the level of cholesterol by 28 points or more.

[5] We make the large-sample assumption and take the sample frequencies as representing $P(y, x|z)$. To account for sample variability, all bounds should be supplemented with confidence intervals and significance levels, as in traditional analyses of controlled experiments. Section 8.5.1 assesses sample variability using Gibbs sampling.

The impact of treatment "on the treated" is equally revealing. Using equation (8.20), $\mathrm{ACE}^*(X \to Y)$ can be evaluated precisely (since $P(x_1|z_0) = 0$):

$$\mathrm{ACE}^*(X \to Y) = \frac{0.465}{0.610} = 0.762.$$

In other words, those subjects who stayed in the program are much better off than they would have been if not treated: the treatment can be credited with reducing cholesterol levels by at least 28 units in 76.2% of these subjects.

8.3 Counterfactuals and Legal Responsibility

Evaluation of counterfactual probabilities could be enlightening in some legal cases in which a plaintiff claims that a defendant's actions were responsible for the plaintiff's misfortune. Improper rulings can easily be issued without an adequate treatment of counterfactuals. Consider the following hypothetical and fictitious case study, specially crafted in Balke and Pearl (1994a) to accentuate the disparity between causal effects and causal attribution.

The marketer of PeptAid (antacid medication) randomly mailed out product samples to 10% of the households in the city of Stress, California. In a follow-up study, researchers determined for each individual whether they received the PeptAid sample, whether they consumed PeptAid, and whether they developed peptic ulcers in the following month.

The causal structure for this scenario is identical to the partial compliance model given by Figure 8.1, where z_1 asserts that PeptAid was received from the marketer, x_1 asserts that PeptAid was consumed, and y_1 asserts that peptic ulceration occurred. The data showed the following distribution:

$$\begin{aligned}
P(y_0, x_0|z_0) &= 0.32, & P(y_0, x_0|z_1) &= 0.02,\\
P(y_0, x_1|z_0) &= 0.32, & P(y_0, x_1|z_1) &= 0.17,\\
P(y_1, x_0|z_0) &= 0.04, & P(y_1, x_0|z_1) &= 0.67,\\
P(y_1, x_1|z_0) &= 0.32, & P(y_1, x_1|z_1) &= 0.14.
\end{aligned}$$

These data indicate a high correlation between those who consumed PeptAid and those who developed peptic ulcers:

$$P(y_1|x_1) = 0.50, \qquad P(y_1|x_0) = 0.26.$$

In addition, the intent-to-treat analysis showed that those individuals who received the PeptAid samples had a 45% greater chance of developing peptic ulcers:

$$P(y_1|z_1) = 0.81, \qquad P(y_1|z_0) = 0.36.$$

The plaintiff (Mr. Smith), having heard of the study, litigated against both the marketing firm and the PeptAid producer. The plaintiff's attorney argued against the producer, claiming that the consumption of PeptAid triggered his client's ulcer and resulting medical expenses. Likewise, the plaintiff's attorney argued against the marketer, claiming that his client would not have developed an ulcer if the marketer had not distributed the product samples.

The defense attorney, representing both the manufacturer and marketer of PeptAid, rebutted this argument, stating that the high correlation between PeptAid consumption and ulcers was attributable to a common factor, namely, pre-ulcer discomfort. Individuals with gastrointestinal discomfort would be much more likely both to use PeptAid and to develop stomach ulcers. To bolster his clients' claims, the defense attorney introduced expert analysis of the data showing that, on average, consumption of PeptAid actually decreases an individual's chances of developing ulcers by at least 15%.

Indeed, the application of (8.14ab) results in the following bounds on the average causal effect of PeptAid consumption on peptic ulceration:

$$-0.23 \leq \text{ACE}(X \rightarrow Y) \leq -0.15;$$

this proves that PeptAid is beneficial to the population as a whole.

The plaintiff's attorney, though, stressed the distinction between the average treatment effects for the entire population and for the subpopulation consisting of those individuals who, like his client, received the PeptAid sample, consumed it, and then developed ulcers. Analysis of the population data indicated that, had PeptAid not been distributed, Mr. Smith would have had at most a 7% chance of developing ulcers—regardless of any confounding factors such as pre-ulcer pain. Likewise, if Mr. Smith had not consumed PeptAid, he would have had at most a 7% chance of developing ulcers.

The damaging statistics against the marketer are obtained by evaluating the bounds on the counterfactual probability that the plaintiff would have developed a peptic ulcer if he had not received the PeptAid sample, given that he in fact received the sample PeptAid, consumed the PeptAid, and developed peptic ulcers. This probability may be written in terms of the parameters q_{13}, q_{31}, and q_{33} as

$$P(Y_{z_0} = y_1 | y_1, x_1, z_1) \quad = \quad \frac{P(r_z{=}1)(q_{13} + q_{31} + q_{33})}{P(y_1, x_1, z_1)},$$

since only the combinations $\{r_x = 1,\ r_y = 3\}$, $\{r_x = 3,\ r_y = 1\}$, and $\{r_x = 3,\ r_y = 3\}$ satisfy the joint event $\{X = x_1,\ Y = y_1,\ Y_{z_0} = y_1\}$. Therefore,

$$P(Y_{z_0} = y_1 | y_1, x_1, z_1) \quad = \quad \frac{q_{13} + q_{31} + q_{33}}{P(y_1, x_1 | z_1)}.$$

This expression is linear in the q parameters and may be bounded using linear programming to give

$$P(Y_{z_0} = y_1 | z_1, x_1, y_1) \geq \frac{1}{p_{11.1}} \max \left\{ \begin{array}{c} 0 \\ p_{11.1} - p_{00.0} \\ p_{11.0} - p_{00.1} - p_{10.1} \\ p_{10.0} - p_{01.1} - p_{10.1} \end{array} \right\},$$

$$P(Y_{z_0} = y_1 | z_1, x_1, y_1) \leq \frac{1}{p_{11.1}} \min \left\{ \begin{array}{c} p_{11.1} \\ p_{10.0} + p_{11.0} \\ 1 - p_{00.0} - p_{10.1} \end{array} \right\}.$$

Similarly, the damaging evidence against PeptAid's producer is obtained by evaluating the bounds on the counterfactual probability

$$P(Y_{x_0} = y_1 | y_1, x_1, z_1) \quad = \quad \frac{q_{13} + q_{33}}{p_{11.1}}.$$

If we minimize and maximize the numerator (subject to (8.13), we obtain

$$P(Y_{x_0} = y_1 | y_1, x_1, z_1) \geq \frac{1}{p_{11.1}} \max \left\{ \begin{array}{c} 0 \\ p_{11.1} - p_{00.0} - p_{11.0} \\ p_{10.0} - p_{01.1} - p_{10.1} \end{array} \right\},$$

$$P(Y_{x_0} = y_1 | y_1, x_1, z_1) \leq \frac{1}{p_{11.1}} \min \left\{ \begin{array}{c} p_{11.1} \\ p_{10.0} + p_{11.0} \\ 1 - p_{00.0} - p_{10.1} \end{array} \right\}.$$

Substituting the observed distribution $P(y, x|z)$ into these formulas, the following bounds were obtained:

$$0.93 \leq P(Y_{z_0} = y_1 | z_1, x_1, y_1) \leq 1.00,$$
$$0.93 \leq P(Y_{x_0} = y_1 | z_1, x_1, y_1) \leq 1.00.$$

Thus, at least 93% of the people in the plaintiff's category would not have developed ulcers had they not been encouraged to take PeptAid (z_0) or, similarly, had they not taken PeptAid (x_0). This lends very strong support for the plaintiff's claim that he was adversely affected by the marketer and producer's actions and product.

In Chapter 9 we will continue the analysis of causal attribution in specific events, and we will establish conditions under which the probability of correct attribution can be identified from both experimental and nonexperimental data.

8.4 A Test for Instruments

As defined in Section 8.2, our model of imperfect experiment rests on two assumptions: Z is randomized, and Z has no side effect on Y. These two assumptions imply that Z is independent of U, a condition that economists call "exogeneity" and which qualifies Z as an instrumental variable (see Sections 5.4.3 and 7.4.5) relative to the relation between X and Y. For a long time, experimental verification of whether a variable Z is exogenous or instrumental has been thought to be impossible (Imbens and Angrist 1994), since the definition involves unobservable factors (or disturbances, as they are usually called) such as those represented by U.[6] The notion of exogeneity, like that of causation itself, has been viewed as a product of subjective modeling judgment, exempt from the scrutiny of nonexperimental data.

The bounds presented in (8.14ab) tell a different story. Despite its elusive nature, exogeneity can be given an empirical test. The test is not guaranteed to detect all violations of exogeneity, but it can, in certain circumstances, screen out very bad would-be instruments.

By insisting that each upper bound in (8.14b) be higher than the corresponding lower bound in (8.14a) we obtain the following testable constraints on the observed distribution:

$$\begin{aligned}
P(y_0, x_0|z_0) + P(y_1, x_0|z_1) &\leq 1, \\
P(y_0, x_1|z_0) + P(y_1, x_1|z_1) &\leq 1, \\
P(y_1, x_0|z_0) + P(y_0, x_0|z_1) &\leq 1, \\
P(y_1, x_1|z_0) + P(y_0, x_1|z_1) &\leq 1.
\end{aligned} \tag{8.21}$$

If any of these inequalities is violated, the investigator can deduce that at least one of the assumptions underlying our model is violated as well. If the assignment is carefully randomized, then any violation of these inequalities must be attributed to some direct influence that the assignment process has on subjects' responses (e.g., a traumatic experience). Alternatively, if direct effects of Z on Y can be eliminated—say, through an effective use of a placebo—then any observed violation of the inequalities can safely be attributed to spurious correlation between Z and U: namely, to assignment bias and hence loss of exogeneity.

The Instrumental Inequality

The inequalities in (8.21), when generalized to multivalued variables, assume the form

$$\max_x \sum_y [\max_z P(y, x|z)] \leq 1, \tag{8.22}$$

which is called the *instrumental inequality.* A proof is given in Pearl (1995b,c). Extending the instrumental inequality to the case where

[6] The tests developed by economists (Wu 1973) merely compare estimates based on two or more instruments and, in case of discrepency, do not tell us objectively which estimate is incorrect.

Z or Y is continuous presents no special difficulty. If $f(y|x,z)$ is the conditional density function of Y given X and Z, then the inequality becomes

$$\int_y \max_z [f(y|x,z)P(x|z)]dy \leq 1 \quad \forall x. \tag{8.23}$$

However, the transition to a continuous X signals a drastic change in behavior, and it seems that the structure of Figure 8.1 induces no constraint whatsoever on the observed density (Pearl 1995c).

From (8.21) we see that the instrumental inequality is violated when the controlling instrument Z manages to produce significant changes in the response variable Y while the treatment X remains constant. Although such changes could in principle be explained by strong correlations between U, X, and Y (since X does not screen off Z from Y), the instrumental inequality sets a limit on the magnitude of the changes.

The similarity of the instrumental inequality to Bell's inequality in quantum physics (Suppes 1988; Cushing and McMullin 1989) is not accidental; both inequalities delineate a class of observed correlations that cannot be explained by hypothesizing latent common causes. The instrumental inequality can, in a sense, be viewed as a generalization of Bell's inequality for cases where direct causal connection is permitted to operate between the correlated observables, X and Y.

The instrumental inequality can be tightened appreciably if we are willing to make additional assumptions about subjects' behavior—for example, that no individual can be discouraged by the encouragement instrument or (mathematically) that, for all u, we have

$$P(x_1|z_1,u) \geq P(x_1|z_0,u).$$

Such an assumption amounts to having no contrarians in the population, that is, no subjects who will consistently choose treatment contrary to their assignment. Under this assumption, the inequalities in (8.21) can be tightened (Balke and Pearl 1997) to yield

$$\begin{aligned} P(y,x_1|z_1) &\geq P(y,x_1|z_0) \\ P(y,x_0|z_0) &\geq P(y,x_0|z_1) \end{aligned} \tag{8.24}$$

for all $y \in \{y_0, y_1\}$. Violation of these inequalities now means either selection bias or direct effect of Z on Y or the presence of defiant subjects.

8.5 Causal Inference From Finite Samples

8.5.1 Gibbs Sampling

This section describes a method of estimating causal effects and counterfactual probabilities from a finite sample, as presented in Chickering and Pearl (1997).[7] The method is applicable within the Bayesian framework, according to which (i) any unknown statistical parameter can be assigned prior probability and (ii) the estimation of that parameter amounts to computing its posterior distribution, conditioned on the sampled data. In our case the parameter in question is the probability $P(r_x, r_y)$ (or $P(r)$ for short), from which we can deduce $\text{ACE}(X \to Y)$.

If we think of $P(r)$ not as probability but rather as the fraction ν_r of individuals in the population who possess response characteristics given by $R = r$, then the idea of assigning probability to such a quantity would fit the standard philosophy of Bayesian analysis; ν_r is a potentially measurable (albeit unknown) physical quantity and can therefore admit a prior probability, one that encodes our uncertainty in that quantity.

Assume there are m subjects in the experiment. We use z^i, x^i, y^i to denote the observed value of Z, X, Y, respectively, for subject i. Similarly, we use r^i to denote the (unobserved) compliance (r_x) and response (r_y) combination for subject i. We use $\mathcal{X}^i$ to denote the triple $\{z^i, x^i, y^i\}$.

Given the observed data $\mathcal{X}$ from the experiment and a prior distribution over the unknown fractions ν_r, our problem is to derive the posterior distribution for $\text{ACE}(X \to Y)$. The posterior distributions of both ν_R and $\text{ACE}(X \to Y)$ can be derived using the graphical model shown in Figure 8.4, which explicitly represents the independencies that hold in the joint (Bayesian) distribution defined over the variables $\{\mathcal{X}, \nu_R, \text{ACE}(X \to Y)\}$. The model can be understood as m realiza-

[7] A similar method, though lacking the graphical perspective, is presented in Imbens and Rubin (1997).

tions of the response-variable model (Figure 8.3), one for each triple in $\mathcal{X}$, connected together using the node representing the unknown fractions $\nu_R = (\nu_{r_1}, \nu_{r_2}, \ldots, \nu_{r_{16}})$. The model explicitly represents the assumption that, given the fractions ν_R, the probability of a subject belonging to any of the 16 compliance-response subpopulations does not depend on the compliance and response behavior of the other subjects in the experiment. From (8.10), ACE$(X \rightarrow Y)$ is a deterministic function of ν_R and consequently ACE$(X \rightarrow Y)$ is independent of all other variables in the domain once these fractions are known.

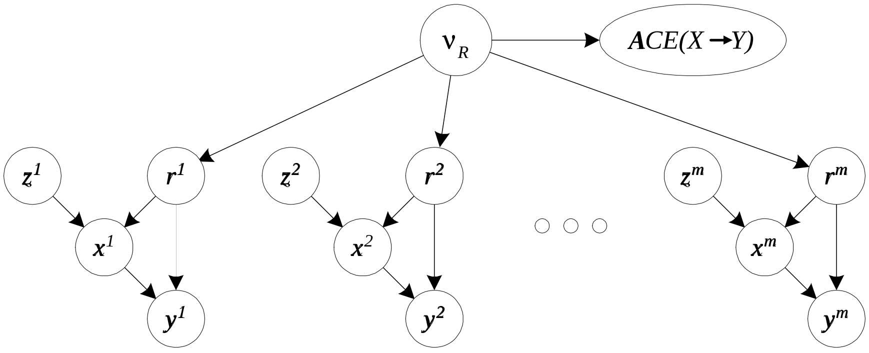


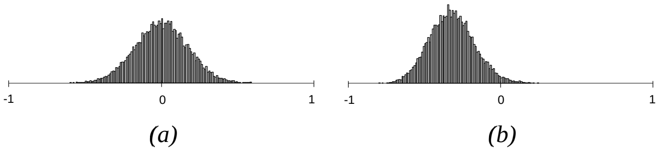
Figure 8.4: Model used to represent the independencies in $P(\{\mathcal{X}\} \cup \{\nu_R\} \cup \{\text{ACE}(X \rightarrow Y)\})$.

In principle, then, estimating ACE$(X \rightarrow Y)$ reduces to the standard inference task of computing the posterior probability for a variable in a fully specified Bayesian network. (The graphical techniques for this inferential computation are briefly summarized in Section 1.2.4.) In many cases, the independencies embodied in the graph can be exploited to render the inference task efficient. Unfortunately, because the r^i are never observed, deriving the posterior distribution for ACE$(X \rightarrow Y)$ is not tractable in our model, even with the given independencies. To obtain an estimate of the posterior distribution of ACE$(X \rightarrow Y)$, an approximation technique known as Gibbs sampling can be used. A graphical version of this technique, called "stochastic simulation," is described in Pearl (1988b, p. 210); the details (as applied to the graph of Figure 8.4) are discussed in Chickering and Pearl (1997). Here we present typical results, in the form of histograms, that demonstrate the general applicability of this technique to problems of causal inference.

8.5.2 The Effects of Sample Size and Prior Distribution

The method takes as input (1) the observed data $\mathcal{X}$, expressed as the number of cases observed for each of the 8 possible realizations of $\{z, x, y\}$, and (2) a Dirichlet prior over the unknown fractions ν_R, expressed in terms of 16 parameters. The system outputs the posterior distribution of ACE$(X \to Y)$, expressed in a histogram.

To show the effect of the prior distribution on the output, we present all the results using two different priors. The first is a flat (uniform) distribution over the 16-vector ν_R, and is commonly used to express ignorance about the domain. The second prior is skewed to represent a strong dependency between the compliance and response characteristics of the subjects. Figure 8.5 shows the distribution of ACE$(X \to Y)$ induced by these two prior distributions (in the absence of any data). We see that the skewed prior of Figure 8.5(b) assigns almost all the weight to negative values of ACE$(X \to Y)$.



(a)

(b)

Figure 8.5: (a) The prior distribution of ACE$(X \to Y)$ induced by flat priors over the parameters ν_{CR}. (b) The distribution for ACE$(X \to Y)$ induced by skewed priors over the parameters.

To illustrate how increasing sample size washes away the effect of the prior distribution, we apply the method to simulated data drawn from a distribution $P(x, y|z)$ for which ACE is known to be identified. Such a distribution is shown Table 8.1. For this distribution, the resulting upper and lower bounds of (8.14ab) collapse to a single point: ACE$(X \to Y)$ $= 0.55$.

Figure 8.6 shows the output of the Gibbs sampler when applied to data sets of various sizes drawn from the distribution shown in Table

z	x	y	$P(x, y, z)$
0	0	0	0.275
0	0	1	0.0
0	1	0	0.225
0	1	1	0.0
1	0	0	0.225
1	0	1	0.0
1	1	0	0.0
1	1	1	0.275

Table 8.1: *Distribution resulting in an identifiable ACE($X \rightarrow Y$)*

8.1, using both the flat and the skewed prior. As expected, as the number of cases increases, the posterior distributions become increasingly concentrated near the value 0.55. In general, because the skewed prior for ACE($X \rightarrow Y$) is concentrated further from 0.55 than the uniform prior, more cases are needed before the posterior distribution converges to the value 0.55.

8.5.3 Causal effects from clinical data with imperfect compliance

In this section we analyze two clinical data sets obtained under conditions of imperfect compliance. Consider first the Lipid Research Clinics Coronary Primary Prevention data described in Section 8.2.6. The resulting data set (after thresholding) is shown in Table 8.2. Using the large-sample assumption, (8.14ab) gives the bounds $0.39 \leq$ ACE($X \rightarrow Y$) ≤ 0.78.

Figure 8.7 shows posterior densities for ACE($X \rightarrow Y$) , based on these data. Rather remarkably, even with only 337 cases in the data set, both posterior distributions are highly concentrated within the large-sample bounds of 0.39 and 0.78.

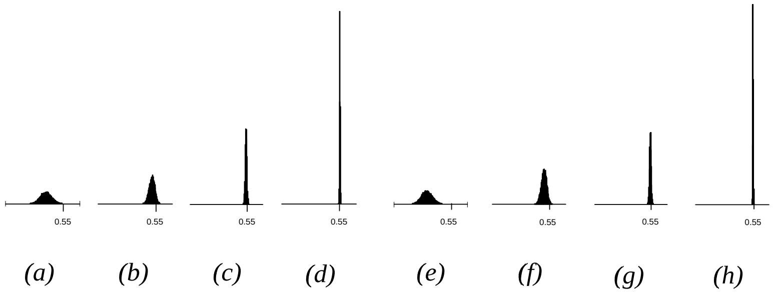


Figure 8.6: Output histograms for identified treatment effect using two priors. (a), (b), (c), and (d) show the posteriors for ACE$(X \rightarrow Y)$ using the flat prior and data sets that consisted of 10, 100, 1,000 and 10,000 subjects, respectively; (e), (f), (g), and (h) show the posteriors for ACE$(X \rightarrow Y)$ using the skewed prior with the same respective data sets.

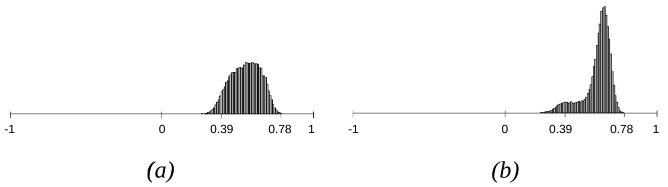


Figure 8.7: Output histograms for the Lipid data. (a) Using flat priors and (b) using skewed priors.

As a second example, we consider an experiment described by Sommer et al. (1986) that was designed to determine the impact of vitamin A supplementation on childhood mortality. In the study, 450 villages in northern Sumatra were randomly assigned to participate in a vitamin A supplementation scheme or serve as a control group for one year. Children in the treatment group received two large doses of vitamin A (x_1), while those in the control group received no treatment (x_0). After the year had expired, the number of deaths y_0 were counted for both groups. The results of this study are also shown in Table 8.2.

Under the large-sample assumption, the inequalities of (8.14ab) yield the bounds $-0.19 \leq$ ACE$(X \rightarrow Y)$ ≤ 0.01. Figure 8.8 shows posterior densities for ACE$(X \rightarrow Y)$, given the data, for two priors.

			Lipid Study	Vitamin A Study
z	x	y	Observations	Observations
0	0	0	158	74
0	0	1	14	11,514
0	1	0	0	0
0	1	1	0	0
1	0	0	52	34
1	0	1	12	2,385
1	1	0	23	12
1	1	1	78	9,663

Table 8.2: *Observed data for the Lipid study and the Vitamin A study*

It is interesting to note that, for this study, the choice of the prior distribution has a significant effect on the posterior. This suggests that if the clinician is not very confident in the prior then a sensitivity analysis should be performed. In such cases, the asymptotic bounds are more informative than the Bayesian estimates, and the major role of the Gibb's sampler would be to give an indication of the sharpness of the boundaries around those bounds.

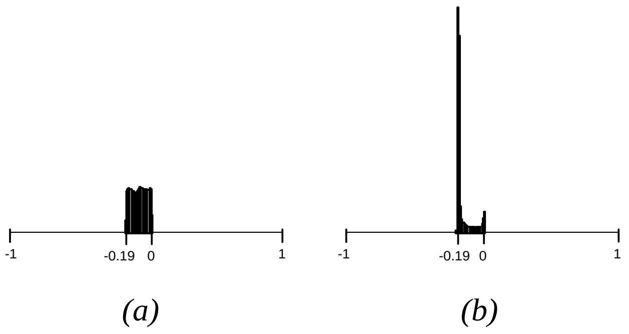


Figure 8.8: Output histograms for the Vitamin A Supplementation data: (a) using flat priors; (b) using skewed priors.

8.5.4 Bayesian Estimate of Single-Event Causation

In addition to assessing causal effects, the Bayesian method just described is also capable (with only minor modification) of answering a variety of counterfactual queries concerning individuals with specific characteristics. Queries of this type were analyzed and bounded in Section 8.3 under the large sample assumption. In this section, we demonstrate a Bayesian analysis of the following query. What is the probability that Joe would have had an improved cholesterol reading had he taken cholestyramine, given that: (1) Joe was in the control group of the Lipid study; (2) Joe took the placebo as prescribed, and (3) Joe's cholesterol level did not improve.

This query can be answered by running the Gibbs sampler on a model identical to that shown in Figure 8.4, except that the function ACE$(X \rightarrow Y)$ (equation (8.10)) is replaced by another function of ν_R, one that represents our query. If Joe was in the control group and took the placebo, that means he is either a complier or a never-taker. Furthermore, because Joe's cholesterol level did not improve, Joe's response behavior is either never-recover or helped. Consequently, Joe must be a member of one of the following four compliance-response populations: $\{(r_x = 0, r_y = 1),\ (r_x = 0, r_y = 2),\ (r_x = 1, r_y = 1),\ (r_x = 1, r_y = 2)\}$. Joe would have improved had he taken cholestyramine if his response behavior is either helped $(r_y = 1)$ or always-recover $(r_y = 3)$. It follows that the query of interest is captured by the function

$$f(\nu_R) = \frac{\nu_{01} + \nu_{11}}{\nu_{01} + \nu_{02} + \nu_{11} + \nu_{12}}.$$

Figures 8.9(a) and (b) show the prior distribution of $f(\nu_R)$ that follows from the flat prior and the skewed prior, respectively. Figures 8.9(c) and (d) show the posterior distribution $P(f(\nu_R|\mathcal{X}))$ obtained from the Lipid data when using the flat prior and the skewed prior, respectively. For reference, the bounds computed under the large-sample assumption are $0.51 \leq f(\nu_R|\mathcal{X}) \leq 0.86$.

Thus, despite 39% noncompliance in the treatment group and despite having just 337 subjects, the study strongly supports the conclusion that—given his specific history—Joe would have been better off

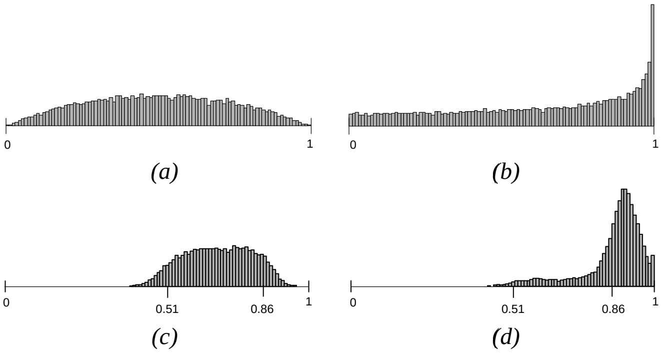


Figure 8.9: Prior (a, b) and posterior (c, d) distributions for a subpopulation $f(\nu_R)$ specified by the counterfactual query: "Would Joe have improved had he taken the drug, given that he did not improve without it." Part (a) corresponds to the flat prior, (b) to the skewed prior.

taking the drug. Moreover, the conclusion holds for both priors.

8.6 Conclusion

This chapter has developed causal-analytic techniques for managing one of the major problems in clinical experiments: the assessment of treatment efficacy in the face of imperfect compliance. Estimates based solely on intent-to-treat analysis—as well as those based on instrumental variable formulas—can be misleading in that they may lie entirely outside the theoretical bounds. The formulas established in this chapter provide instrument-independent guarantees for policy analysis and, in addition, should enable analysts to determine the extent to which efforts to enforce compliance may increase overall treatment effectiveness.

The importance of indirect experimentation is not confined to studies involving human subjects. Experimental conditions equivalent to those of imperfect compliance occur whenever the variable whose causal effect we seek to assess cannot be manipulated directly yet could be partially influenced by indirect means. Typical applications involve the diagnosis of ongoing processes for which the source of malfunctioning behavior must be identified using indirect means because direct manipulation of suspected sources is either physically impossible or prohibitively expensive. An example of the latter would be interrupting the normal operation of a production line so as to achieve direct control over a physical parameter that is suspected of malfunctioning. Partial control over that parameter, in the form of indirect influence, would be much more convenient and would allow the production to continue.

Methodologically, the message of this chapter has been to demonstrate that, even in cases where causal quantities are not identifiable, reasonable assumptions about the salient relationships in the domain can be harnessed to yield useful quantitative information about the causal forces that operate in the domain. Once such assumptions are articulated in graphical form, they can easily be submitted to algebraic methods that yield the desired bounds or, alternatively, invite Gibbs sampling technique to facilitate Bayesian estimation of the causal quantities of interest.

Acknowledgment

The main results in this chapter are based on collaborative works with Alex Balke and David Chickering. The encouragements of James Robins and Miles Hollander helped us endure a long and agonizing publication process. These investigations have benefited from discussions with Phil Dawid, James Heckman, Guido Imbens, Steve Jacobsen, Steffen Lauritzen and Charles Manski.