11.3.2 Demystifying "Strong Ignorability"

Researchers working within the confines of the potential-outcome language express the condition of "zero bias" or "no-confounding" using an independence relationship called

[3] Recall that Greenland and Robins (1986) were a lone beacon of truth for many years, and even they had to resort to the "black-box" language of "exchangeability" to define "bias," which discouraged intuitive interpretations of confounding (see Section 6.5.3). Indeed, it took epidemiologists another six years (Weinberg 1993) to discover that adjusting for factors affected by the exposure (as in Figure 11.5) would introduce bias.

"strong ignorability" (Rosenbaum and Rubin 1983). Formally, if X is a binary treatment (or action), strong ignorability is written as:

$$\{Y(0), Y(1)\} \perp\!\!\!\perp X \mid Z, \tag{11.4}$$

where $Y(0)$ and $Y(1)$ are the (unobservable) potential outcomes under actions $do(X = 0)$ and $do(X = 1)$, respectively (see equation (3.51) for definition), and Z is a set of measured covariates. When "strong ignorability" holds, Z is *admissible,* or *deconfounding*, that is, treatment effects can be estimated without bias using the adjustment estimand, as shown in the derivation of equation (3.54).

Strong ignorability, as the derivation shows, is a convenient syntactic tool for manipulating counterfactual formulas, as well as a convenient way of formally assuming admissibility (of Z) without having to justify it. However, as we have noted several times in this book, hardly anyone knows how to apply it in practice, because the counterfactual variables $Y(0)$ and $Y(1)$ are unobservable, and scientific knowledge is not stored in a form that allows reliable judgment about conditional independence of counterfactuals. It is not surprising, therefore, that "strong ignorability" is used almost exclusively as a surrogate for the assumption *"Z* is admissible," that is,

$$P(y \mid do(x)) = \sum_z P(y \mid z, x)P(z), \tag{11.5}$$

and rarely, if ever, as a criterion to protect us from bad choices of Z.[4]

Readers enlightened by graphical models would recognize immediately that equation (11.4) must mirror the back-door criterion (p. 79, Definition 3.3.1), since the latter too entails admissibility. This recognition allows us not merely to pose equation (11.4) as a claim, or an assumption, but also to reason about the cause–effect relationships that render it valid.

The question arises, however, whether the variables $Y(0)$ and $Y(1)$ could be represented in the causal graph in a way that would allow us to test equation (11.4) by graphical means, using d-separation. In other words, we seek a set W of nodes such that Z would d-separate X from W if and only if Z satisfies equation (11.4).

The answer follows directly from the rules of translation between graphs and potential outcome (Section 3.6.3). According to this translation, $\{Y(0), Y(1)\}$ represents the sum total of all exogenous variables, latent as well as observed, which can influence Y through paths that avoid X. The reason is as follows: according to the structural definition of $\{Y(0), Y(1)\}$ (equation (3.51)), $Y(0)$ (similarly $Y(1)$) represents the value of Y under a condition where all arrows entering X are severed, and X is held constant at $X = 0$. Statistical variations of $Y(0)$ would therefore be governed by all exogenous ancestors of Y in the mutilated graphs with the arrows entering X removed.

In Figure 11.4, for example, $\{Y(0), Y(1)\}$ will be represented by the exogenous variables $\{e_1, e_2, e_3, e_4\}$. In Figure 3.4, as another example, $\{Y(0), Y(1)\}$ will be represented by the noise factors (not shown in the graph) that affect variables X_4, X_1, X_2, X_5, and X_6. However,

[4] In fact, in the rare cases where "strong ignorability" is used to guide the choice of covariates, the guidelines issued are wrong or inaccurate, perpetuating myths such as: "there is no reason to avoid adjustment for a variable describing subjects before treatment," "a confounder is any variable associated with both treatment and disease," and "strong ignorability requires measurement of all covariates related to both treatment and outcome" (citations withheld to spare embarrassment).

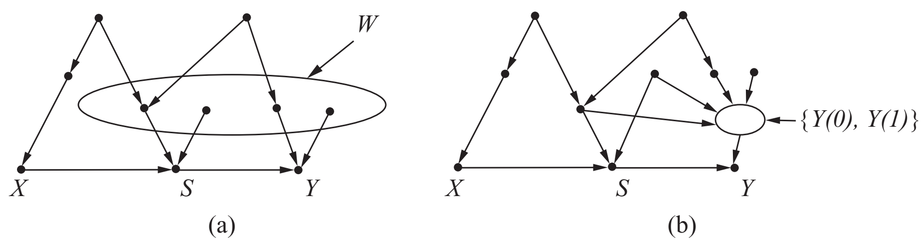

Figure 11.7 Graphical interpretation of counterfactuals $\{Y(0), Y(1)\}$ in the "strong ignorability" condition.

since variables X_4 and X_5 summarize (for Y) the variations of their ancestors, a sufficient set for representing $\{Y(0), Y(1)\}$ would be X_4, X_1 and the noise factors affecting Y and X_6.

In summary, the potential outcomes $\{Y(0), Y(1)\}$ are represented by the observed and unobserved parents[5] of all nodes on paths from X to Y. Schematically, we can represent these parents as in Figure 11.7(a). It is easy to see that, with this interpretation of $\{Y(0), Y(1)\}$, a set of covariates Z d-separates W from X if and only if Z satisfies the back-door criterion.

It should be noted that the set of observable variables designated W in Figure 11.7(a) are merely surrogates of the unobservable counterfactuals $\{Y(0), Y(1)\}$ for the purpose of confirming conditional independencies (e.g., equation (11.4)) in the causal graph (via d-separation.) A more accurate allocation of $\{Y(0), Y(1)\}$ is given in Figure 11.7(b), where they are shown as (dummy) parents of Y that are functions of, though not identical to, the actual (observable) parents of Y and S.

Readers versed in structural equation modeling would recognize the graphical representations $\{Y(0), Y(1)\}$ as a refinement of the classical economentric notion of "disturbance," or "error term" (in the equation for Y), and "strong ignorability" as the requirement that, for X to be "exogenous," it must be independent of this "disturbance" (see Section 5.4.3). This notion fell into ill repute in the 1970s (Richard 1980) together with the causal interpretation of econometric equations, and I have predicted its re-acceptance (p. 170) in view of the clarity that graphical models shine on the structural equation formalism. Figure 11.7 should further help this acceptance.

Having translated "strong ignorability" into a simple separation condition in a model that encodes substantive process knowledge should demystify the nebulous concept of "strong ignorability" and invite investigators who speak "ignorability" to benefit from its graphical interpretation.

This interpretation permits researchers to understand what conditions covariates must fulfill before they eliminate bias, what to watch for and what to think about when covariates are selected, and what experiments we can do to test, at least partially, if we have the knowledge needed for covariate selection. Section 11.3.4 exemplifies such considerations.

One application where the symbiosis between the graphical and counterfactual frameworks has been useful is in estimating the effect of treatments on the treated: ETT = $P(Y_{x'} = y \mid x)$ (see Sections 8.2.5 and 11.9.1). This counterfactual quantity (e.g., the probability that a treated person would recover if not treated, or the rate of disease among the exposed, had the exposure been avoided) is not easily analyzed in the do-calculus notation. The counterfactual notation, however, allows us to derive a

[5] The reason for explicitly including latent parents is explained in Section 11.3.1.

useful conclusion: Whenever a set of covariates Z exists that satisfies the back-door criterion, ETT can be estimated from observational studies. This follows directly from

$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \Longrightarrow Y_{x'} \perp\!\!\!\perp X \mid Z,$$

which allows us to write

$$\begin{aligned}
\text{ETT} &= P(Y_{x'} = y \mid x) \\
&= \textstyle\sum_z P(Y_{x'} = y \mid x, z) P(z \mid x) \\
&= \textstyle\sum_z P(Y_{x'} = y \mid x', z) P(z \mid x) \\
&= \textstyle\sum_z P(y \mid x', z) P(z \mid x).
\end{aligned}$$

The graphical demystification of "strong ignorability" also helps explain why the probability of causation $P(Y_{x'} = y' \mid x, y)$ and, in fact, any counterfactual expression conditioned on y, would not permit such a derivation and is, in general, non-identifiable (see Chapter 9).