A Review of Causal Inference

by

Dayang Liu

A Research Review Report

Submittted to the faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

by

December 2008

APPROVED:

Professor Joseph Petruccelli, Advisor

Professor Bogdan Vernescu, Department Head

Abstract

In this report, I first review the evolution of ideas of causation as it relates to causal inference. Then I introduce two currently competing perspectives on this issue: the counterfactual perspective and the noncounterfactual perspective. The ideas of two statisticians, Donald B. Rubin, representing the counterfactual perspective, and A.P.Dawid, representing the noncounterfactual perspective are examined in detail and compared with the evolution of ideas of causality. The main difference between these two perspectives is that the counterfactual perspective is based on counterfactuals which cannot be observed even in principle but the noncounterfactual perspective only relies on observables. I describe the definition of causes and causal inference methods under both perspectives, and I illustrate the application of the two types of methods by specific examples. Finally, I explore various controversies on these two perspectives.

Contents

Chapter 1

Introduction

Establishing cause is of vital practical interest in many disciplines. In medicine, it is imperative to know that medications cause the desired effects. In epidemiology, people look for causes of diseases. In social sciences, researchers try to discover the causes of human behaviors.

However, behind the practical need to establish cause is the basic question of what constitutes cause and when one can show there is a cause. Philosophy and statistics are two disciplines that consider such fundamental questions. Cause is of theoretical interest in philosophy and of theoretical and methodological interest in statistics. Philosophers have been discussing the idea of cause for millennia, and continue that discussion today. Causal inference has been explored by statisticians for nearly a century and continues to be an active research area in statistics.

This paper is a research review report about causal inference in statistics. Its aim is to present a survey of some recent research in causal inference. It focuses on one of the most active areas of recent research: causal models involving counterfactuals. These models are based on ideas first broached by statisticians in the early twentieth century, and so the paper also considers the historical development of these ideas. Not all statisticians agree with the validity of basing inference on counterfactuals, which are unobservable in principle as well as in practice. Consequently, I present

some competing views of what constitutes valid causal inference.

The report is based on readings of a number of key papers. I have synthesized the material from those papers and divided it into the following four chapters. In chapter 2, I review some early thoughts on causality from philosophers and statisticians (Holland 1986). It is shown that philosophers have widely different ideas and distinct emphases. But several statisticians seem to implicitly share the same basic perspective. In chapter 3, I introduce Rubin's Counterfactual Model (Rubin 1974, 1978, 1990, 2004) (Holland and Rubin 1983), which has been the focus of much of the recent statistical research in causal inference. I first describe the basic elements of this model. Then I compare the definition of causation in this model with historical definitions and explain the inference methods behind this model. At the end, I give an example of a specific application of this model. In chapter 4, I introduce an important competing view for causal inference, Dawid's Noncounterfactual method (Dawid 2000). I first illustrate Dawid's criticism of Rubin's counterfactual model by a specific example and then describe the idea of Dawid's noncounterfactual perspective and his inference method. In chapter 5, I show some controversies on these two contradictory points of view among statisticians (Dawid 2000). The emphases are put on four topics: positivism, untestable assumptions, instrumental use, and application.

Chapter 2

History

A great deal has been said about causality by philosophers and statisticians, so it is impossible to give a complete coverage of their ideas. This chapter views some important contributions.

2.1 Ideas from philosophers

Aristotle listed four causes of a thing in his physics (Holland 1986): The material cause (that out of which the thing is made), the formal cause (that into which the thing is made), the efficient cause (that which makes the thing), and the final cause (that for which the thing is made). His notion of efficient cause is the one that is close to the usual modern definition of cause.

Locke (1690) introduced his definitions on causality (Holland 1986): " That which produces any simple or complex idea, we denote by the general name 'cause', and that which is produced, 'effect'." It should be noted that Locke's notion included the idea of an effect, which differed from that of Aristotle who only focused on cause.

Hume (1740, 1748) proposed three basic criteria for causation (Holland 1986): (a) spatial/temporal contiguity, (b) temporal succession, and (c) constant conjunction.

For Hume, in order to show A causes B, it is necessary that (a) A and B be contiguous in space and time, (b) A precede B in time, and (c) A and B occur (or do not occur) together.

Mill (1843) provided some ideas regarding how to discover causation in practice (Holland 1986). He described four methods: (a) the method of concomitant variation: if Y varies as A varies, A might be a cause of the change in Y; (b) the method of difference: the difference between Y when A happens and when B happens indicates the cause; (c) the method of residues: the effect of B on Y can be observed by taking the difference between Y when A and B both happen and that when only A happens; and (d) the method of agreement: if Y does not change regardless of A or B happening, neither A nor B cause change in Y.

2.2 Ideas from Statisticians

After reviewing some ideas from philosophers, I will take a look at some early statisticians' thoughts. Although the exact idea was not always stated explicitly, many statisticians brought up the idea of multiple versions of the responses and considered the difference between them to be causal effects.

Kempthorne (1952) in a discussion of the analysis of an agricultural experimental plan (in which larger tracts of land, called blocks, are each subdivided into p plots and then one of the experimental treatments is applied at random to each of the p plots within each block.) defined yields as follows: "We shall denote the yield with treatment k ... on plot j of block I ... by y_{ijk}." He then wrote: "In fact we do not observe the yield of treatment k on plot j but merely the yield of treatment k on a randomly chosen plot in the block ... we denote the observed yield of treatment k in block i by y_{ik}". It seems evident that the y_{ijk} in the first quotation refers to different versions of the response-one for each k-on each combination (i, j) of plot within block. The y_{ik} in the second quotation is the value of y_{ijk} for that plot to which treatment k is actually applied in block i (Holland 1986).

In an experiment with treatments T_1, T_2, D.R.Cox (1958) defined the true treatment effects as the difference between "the observation obtained on any unit when, say, T_1 is applied" and "the observation that would have been observed had, say, T_2 been applied", namely, the difference between multiple versions of responses (Holland 1986).

Earlier, Fisher (1926) and Neyman (1935) had expressed similar ideas, though their disagreement on the nature of the resulting inference caused controversy. Fisher insisted that inference be conducted at the individual unit level. Assume for simplicity that each unit can be given either treatment or"control", which might denote a sham treatment. For Fisher, the null hypothesis of no causal effect was that the difference between the response under the treatment and the response under the control equal zero for each unit. In contrast, Neyman considered the null hypothesis of no treatment effect to be that the difference in mean effect, computed over the entire population, equal zero (Holland 1986).

Chapter 3

Rubin's Counterfactual Model

In statistics, there are two main competing ideas for causal inference currently: the counterfactual perspective and noncounterfactual perspective. One particular model based on the counterfactual perspective that I will explore in detail has been primarily developed by Donald B. Rubin over the last 35 years (Rubin 1974, 1978, 1990, 2004) (Holland 1986). Rubin's model extends ideas of earlier philosophers and statisticians, particularly Kempthorne, D.R.Cox and Neyman (Holland 1986). Generally, a counterfactual is a conditional statement the first clause of which expresses something contrary to fact, as "If I had known". In Rubin's model, counterfactuals play a specific role, as will be shown.

3.1 Model Elements

A unit is a single thing or person, denoted by u. Units can be considered coming from a population U. Without loss of generality, I consider two causes or treatments: treatment (W=1) and control (W=0). Treatments are applied to the units.

Before any treatment is applied, there are two potential responses from unit u: $Y_1(u)$ and $Y_0(u)$. $Y_1(u)$ is the response from unit u when treatment 1 is applied. $Y_0(u)$

is the response from unit u when treatment 0 is applied. The particular time when treatment is applied and when the response is measured must be specified since the same treatment applied at different times might cause different effects. Similarly, responses which are measured at different times are different responses. In Rubin's model, therefore, only one of the two treatments can be applied, and only the response to that treatment can be observed.

Once treatment 1 is applied to u, $Y_0(u)$ becomes a counterfactual since treatment 0 was not applied. As a result, $Y_0(u)$ cannot be observed even in principle. Similarly, if treatment 0 is applied to unit u, $Y_0(u)$ is observed and $Y_1(u)$ is the counterfactual. In other words, one of the potential responses would become the actual response and the other would become the counterfactual.

In some causality analyses, researchers are interested in the difference of the causal effects across different groups of units. As a result, a group variable G, which denotes the group to which a unit belongs, is present in many analyses. Finally, (causal) effect of 1 (relative to 0) on u is defined as $\tau(u) = Y_1(u) - Y_0(u)$, namely, the difference between the response under treatment 1 and that under treatment 0. Note that $\tau(u)$ cannot be observed because it is the difference between an actual response and a counterfactual, and the latter cannot be observed.

Covariates are variables whose values are not affected by the treatment assignment. Covariates are denoted as X. The assignments of treatments often depend on particular covariates, and some covariates might also have causal relationships with the responses. Therefore, it is usually preferable to adjust the inferences for some covariates (Rubin 1974, 1978, 1990, 2004).

A simple example will better illustrate the basic elements in Rubin's counterfactual model. *The subjects are patients who are having a cold. Half of male patients are treated with aspirin and the other half are treated with a new drug. Separately, the same treatment assignment is used for the female patients. For each gender group, will the proportion of patients whose colds disappear in two days differ for those given*

treatment (the new drug) and control (aspirin)? Is there any difference in the results between male and female groups?

In this example, the units are patients. The treatments are aspirin ($W = 0$), and the new drug ($W = 1$). The two potential responses are "cold disappears in two days" ($Y_W(u) = 1$), and "cold does not disappear in two days" ($Y_W(u) = 0$). The group variable is gender. Covariates could be age, race, weight and so on. The effect of the new drug (relative to aspirin) on patient u is $\tau(u) = Y_1(u) - Y_0(u)$.

3.2 What Can Be A Cause?

Before introducing causal inference, it will be helpful to review the old ideas of cause again in the context of Rubin's counterfactual model. Aristotle defined cause without reference to any effect of the cause. Locke gave a general definition of an effect but without the notion of difference between effects of different causes.

Hume's first criterion is expressed in the counterfactual model since the application of a treatment and the measurement of the response take place on a common unit (spatial contiguity). The time period between the measurement of a response and the application of a treatment affects the effect (temporal contiguity). Temporal succession is also embraced by the counterfactual model since the measurement of a response always happens after the application of a treatment. However, constant conjunction might fail since the causal effect $Y_1(u) - Y_0(u)$ could vary with the unit u. Hume didn't have the notion that the effect is always relative to another cause either.

Mills's method of concomitant variation might imply causation, but it can also result from mere correlation. The methods of difference and residues can be regarded as useful for discovering the causal effect in the counterfactual model. And the method of agreement is just for identifying no effect in the model.

The definition of an effect as the difference between multiple versions of responses,

initially proposed by statisticians, is exactly same as the definition in the counterfactual model. Moreover, Fisher and Neyman further proposed methods for causal inference (Rubin 2004).

It should be noted that Rubin's counterfactual model does not address all possible causes, and it only accommodates treatments which can be applied at least in principle. Examples given by Holland (1986) explain this idea in more detail. (A)"*She did well on the exam because she was coached by her teacher.*" (B) "*She did well on the exam because she is a woman.*" (C) "*She did well on the exam because she studied for it.*" In (A), whether a person is coached or not can usually be determined so that they are valid causes and can be applied in practice. In (B), the gender of a person is an attribute of this unit which cannot be manipulated. Even though some extreme situations are considered, the unit itself would be changed if his/her gender were switched. In this sense, gender is not a valid cause in the counterfactual model. In some cases, it is hard to tell if one thing is a cause or not. In (C), willingness to study might be considered an attribute that some people are born with and others are not. And no one can be educated to become more or less willing to study. Moreover, even though a person is inclined to study in general, he/she might not study due to some other uncontrolled causes: for instance, forgetting to bring the book home. On the other hand, some other people might think education certainly affects one's willingness to study. Therefore, the"validity" of a cause sometimes depends on opinion.

Another important consideration about causes in Rubin's counterfactual model is that treatments usually consist of a series of actions and the effects in this instance should be attributed to the entire series of actions. For example, consider the statement "*I took two aspirin and a cup of water, the headache then went away.*" Obviously, taking water and aspirin constitute the treatment in this example and the effect is ascribed to both of them. Usually, for convenience, it is not necessary to state every action in the series. One would say "*I took two aspirin, the headache then went away.*" But knowing what is the real treatment clearly is important.

3.3 Causal Inference

This section considers the use of Rubin's counterfactual model in causal inference. It begins by introducing the Fundamental Problem of Causal Inference (FPCI) and solutions to it. Lord's paradox provides a specific example of causal inference under Rubin's counterfactual model. (Holland and Rubin, 1983)

3.3.1 The Fundamental Problem of Causal Inference

The Fundamental Problem of Causal Inference (FPCI) states that it is impossible to observe the value of $Y_1(u)$ and $Y_0(u)$ on the same unit and, therefore, it is impossible to observe the effect of Treatment 1 (relative to Treatment 0) on u. Upon observation, one of the two potential responses becomes the actual observed response, but the other response remains unobserved: it is the counterfactual that could never be observed even in principle. Therefore, the difference $Y_1(u) - Y_0(u)$, the effect, cannot be observed directly.

However, in certain scientific settings it may be reasonable to assume that the FPCI does not apply. For example, this can occur in the presence of temporal stability and causal transience. Temporal stability means that the response does not change if the time when a treatment is applied is varied slightly. Causal transience means that the response of one treatment is not affected by prior exposure of the unit to the other treatment. If these two assumptions are plausible, one can simply apply one treatment first and then the other and treat them as if both had been applied to the unit at the same time. The difference of the two observed responses is therefore equal to the effect.

Another assumption that is often used to avoid the FPCI is the unit homogeneity assumption, which states that the units are homogeneous with respect to the treatments and responses. If this assumption holds, one can apply two different treatments to two different units respectively. The observed response of one unit is then treated

as the counterfactual of the other unit.

Although these assumptions are widely used in a variety of areas, there are many situations in which none of them is plausible. In such cases, statistical solutions may provide another option.

3.3.2 Assignment Mechanism of Treatments

The assignment mechanism of treatments (causes) plays an essential role in causal inference. Let W denote the vector of treatment assignments for the units. Unconfounded assignment mechanisms are defined as those which satisfy:

$$P_r[W|X, Y_1(U), Y_0(U)] = P_r(W|X).$$

Unconfounded assignment mechanisms do not depend on any responses, but may depend on covariates. Assignment mechanisms are called ignorable if

$$P_r[W|X, Y_1(U), Y_0(U)] = P_r(W|X, Y_{obs})$$

Ignorable assignment mechanisms may depend on observed responses in addition to covariates, but not on counterfactuals. "Ignorable" means that the assignment mechanisms are totally understood since those which cannot be observed, the counterfactuals, have no effect, and can therefore be "ignored". Nonignorable assignment mechanisms are those which rely on counterfactuals. Therefore, one doesn't know how the mechanisms work entirely since counterfactuals cannot be ignored.

Most analyses, make the "Stable-Unit-Treatment-Value Assumption" (SUTVA), meaning that the response of a unit u to treatment W is not affected by what treatments other units receive. For instance, in two different assignments, if unit u receives the same treatment W, then under SUTVA, $Y_W(u)$ would be the same value regardless of how different are the treatment assignments for other units.

The completely randomized treatment assignment mechanism is the simplest ignorable/unconfounded assignment mechanism since all the assignments share an equal

constant probability that doesn't depend on covariates, counterfactuals, or other units. Regular designs are similar to completely randomized treatment assignment mechanism except that the probabilities of treatment assignment are allowed to depend on covariates, and so can vary from unit to unit. Regular designs are the major template for the analysis of experiments and observational studies. They have two properties. First, they use an unconfounded assignment mechanism. Second, $Pr(W|X) = g(W)\prod_1^N p_i$, where $g()$ is an exchangeable function ($g(x_1, \ldots, x_n)$ is an exchangeable function if $g(x_1, \ldots, x_n) = g(P(x_1, \ldots, x_n))$, where $P(.)$ is any permutation of $x_1, \ldots, x_n$.) and $p_i \equiv P_r(W_i|X_i)$, $0 < p_i < 1$. The p_i are usually called propensity scores (Rubin 2000).

3.3.3 Inference Methods

There are three common statistical inference methods useful in causal inference. I first introduce their applications in completely randomized experiments.

Fisherian randomization-based inference is the first method. Suppose there is a population U, and that the observed responses for all units in $u \in U$ are known. The null hypothesis is $\tau(u) = 0$ for all $u \in U$. Under this hypothesis, the counterfactual is equal to the observed response for each unit. Therefore, for each possible assignment, the value of the statistic $S = E(Y(u)|T = 1) - E(Y(u)|T = 0) = \bar{y_1} - \bar{y_0}$ is known. This allows one to conduct an exact test and obtain a p value based on the permutation distribution of S (Rubin 1974). Fisher's approach can be extended to additive null hypotheses (ANH): $\tau(u) = c$ where c is any constant, for all u. The test statistic is calculated by substituting $Y_1(u) - c$ for $Y_1(u)$, and the analysis is then performed in the same way as described above (Rubin 2004).

The second method is Neymanian randomization-based inference. Neymanian inference concerns the average effect on U: $E(\tau(u)) = E(Y_1(u)) - E(Y_0(u))$ instead of the effects at the individual level. Under completely random assignment, an unbiased estimate for $E(\tau)$ is the difference between the average response of the units receiv-

ing treatment 1 and those receiving treatment 0: $E(\tau(u)) = E(Y_1(u)) - E(Y_0(u)) = E(Y(u)|T = 1) - E(Y(u)|T = 0) = \bar{y_1} - \bar{y_0}$. Assuming ANH holds, an unbiased estimate of the variance of τ, $\sigma^2(\tau)$, is $(se)^2 = s_1^2/n_1 + s_0^2/n_0$, where s_1^2 and s_0^2 are the standard deviations of the Treatment 1 and Treatment 0 groups respectively, and n_1 and n_0 are the sizes for the two groups. As a result, a 95% confidence interval for $E(\tau(u))$ can be obtained using normal approximation: $\bar{y_1} - \bar{y_0} \pm 1.96se$.

The last method is Bayesian inference. Here potential responses are treated as random variables. A prior model for them given the covariates is specified: $Pr(Y_1(u), Y_0(u)|X)$. Under this model, the unobservable counterfactuals are assumed to be missing values. Using Bayesian methods, one can obtain a posterior predictive distribution of the counterfactuals Y_{mis}, $Pr(Y_{mis}|W, X, Y_{obs})$, based on the observed responses Y_{obs}, and the prior model. Finally, the predictive distribution of the effects can be easily obtained by taking differences. Due to the reliance of the prior model and the assignment on the covariates, the result can be highly sensitive to the prior distribution when covariates have unbalanced distributions in the two groups (treatment and control).

In controlled experiments, with more complex randomized assignment mechanisms with known propensity scores, the analyses almost proceed as in completely randomized experiments. Although probabilities of assignments are not constants but depend on covariates, they are still known. For Fisherian inference, the distribution of S can be obtained so that the exact test can still be conducted. For Neymanian inference, the results under regular assignment mechanisms are still valid. For Bayesian inference, the prior model is more complicated but the analysis still follows the same method (Rubin 2004).

In observational studies, sometimes one holds a "Strongly Ignorable Treatment Assignment" assumption meaning that assignment mechanisms are still viewed as they are in controlled experiments but with unknown propensity scores. In this situation, many methods (e.g., discriminant analysis, logistic regression) can be used to estimate propensity scores. Once one has the estimates, the analyses are same as

those with known propensity scores. But the "Strongly Ignorable Treatment Assignment" assumption is certainly not plausible in many cases. In such cases, conducting sensitivity analyses to get bounds on estimates is the best one can do. Stratification/Blocking/Matching are common approaches to adjust for propensity scores so that less sensitive, more precise results can be obtained (Rubin 2004).

Covariates that have causal relationships with responses can aid inference. Such covariates are known as "correct covariates". For example, suppose researchers are analyzing the causal relationship between a new drug and death for some seriously ill patients. From previous research, it is known that smoking has a causal relationship with death. Then in this analysis, smoking or not is a correct covariate. Adjusting for correct covariates results in a more sensitive test in the Fisherian method, a shorter interval in the Neymanian method, and less sensitivity to model specification in the Bayesian method. Stratification/Blocking/Matching are still the common methods for adjustment (Rubin 2004).

3.3.4 Lord's Paradox

Lord's paradox (Lord 1967) presents a situation in which two statisticians, using different, but standard, statistical methods come to contradictory conclusions. Here, I show how the paradox can be resolved using Rubin's counterfactual model (Holland and Rubin 1983).

Lord introduces the problem thus: "*A large university is interested in investigating the effects on the students of the diet provided in the university dining halls and any sex differences in these effects. Various types of data are gathered. In particular, the weight of each student at the time of his/her arrival in September and his/her weight the following June are recorded.*"

The data were fabricated by Lord, but for our purposes, the important information is:
The average weight for males was 180 in both September and June.

The average weight for females was 130 in both September and June.
The average weight gain for males was zero.
The average weight gain for females was zero.

In Lord's description, Statistician 1 looks at gain scores and concludes that there was no effect of diet on weight, and no evidence of different effect for the two sexes, as weight gain was zero in both groups. Statistician 2 is interested in weight gain for males and females with the same weight in September, and finds that on average, for a given September weight, men weighed more in June than women. Thus, the new diet resulted in more weight gain for men. The two conclusions are obviously contradictory to each other. However, both of them are based on the same data and on standard statistical methods, and seem to be reasonable. This contradiction is the source of Lord's paradox.

In order to use the counterfactual model, one needs to first identify the basic elements in this example. The results are shown in Table 3.1:

Table 3.1:

Study Design	
U:	The students at the university in the specified school year.
Treatment 1:	The dining hall diet,
Treatment 0:	?
W:	$W = 1$ for all units.
Variable Measured	
G:	Student gender (1=male, 2=female),
X:	The weight of a student in September,
Y:	The weight of a student in June.

The question mark in Table 3.1 is due to no control diet, 0, even hinted at in the example. Since no one is exposed to 0, one is forced to make untestable assumptions on $Y_0(u)$ in order to obtain an estimate of the causal effect. The analyses are summarized in Table 3.2.

Table 3.2:

	Statistican 1
Testable Assumptions	-
Untestable Assumptions	$\alpha = 0, \beta = 1$
Formula for Causal Effects Δ_i	$\Delta_i = E(Y_1 - X \mid G = i)$
Formula for Differential Causal Effect Δ	$\Delta = E(Y_1 \mid G = 1) - E(Y_1 \mid G = 2) - [E(X \mid G = 1) - E(X \mid G = 2)]$ =difference in mean weight gains
Assume $Y_0 = \alpha + \beta X$ for all units in U.	
	Statistican 2
Testable Assumptions	$E(Y_1 \mid X, G = i) = a_i + bX$
Untestable Assumptions	$\beta = b$
Formula for Causal Effects Δ_i	$\Delta_i = E(Y_1 - \alpha - bX \mid G = i)$
Formula for Differential Causal Effect Δ	$\Delta = E(Y_1 \mid G = 1) - E(Y_1 \mid G = 2) - b[E(X \mid G = 1) - E(X \mid G = 2)]$ =covariate adjusted mean difference in June weights.
Assume $Y_0 = \alpha + \beta X$ for all units in U.	

In the analyses which are associated with the ideas of the two statisticians, one assumes $Y_0 = \alpha + \beta X$ for all units in U. If one makes an untestable assumption that $\alpha = 0, \beta = 1$, the average effect of 1 in each group would be $\Delta_i = E(Y_1 - Y_0 | G = i) = E(Y_1 - X | G = i)$. Then the difference between the effects of the two gender groups is $\Delta = \Delta_1 - \Delta_2 = E(Y_1 | G = 1) - E(Y_1 | G = 2) - [E(X | G = 1) - E(X | G = 2)]$ which is the difference in mean weight gains estimated by Statistician 1. As mentioned above, it is zero so that there is no difference of the effects of 1 between males and females. However, if one makes a testable assumption that $E(Y_1 | X, G = i) = a_i + bX$ and an untestable assumption $\beta = b$, the differences of the effects would be $\Delta = \Delta_1 - \Delta_2 = E(Y_1 - Y_0 | G = 1) - E(Y_1 - Y_0 | G = 2) = E(Y_1 | G = 1) - E(Y_1 | G = 2) - b[E(X | G = 1) - E(X | G = 2)]$ which is the covariate-adjusted mean difference in June weights estimated by Statistician 2.

In summary, Rubin's counterfactual model reveals that, although based on the same data, the two contradictory conclusions actually come from two sets of different assumptions. The contribution of the counterfactual model is to make clear what those unstated assumptions are. That different assumptions lead to distinct conclusions

should not be a surprise. However, the paradox arises not solely because the assumptions differ, but also because they are untestable from the data, as the analysis with the counterfactual model makes clear.

Chapter 4

The Noncounterfactual Pespective

Some statisticians think that causal inferences should not be based on unobservable counterfactuals. They believe that appropriate analysis methods should only rely on observables and testable assumptions. I call this view the Noncounterfactual Perspective. A. P. Dawid is a leading proponent of the Noncounterfactual Perspective. In the following sections, I introduce Dawid's ideas on causal inference, his criticisms on Rubin's counterfactual model, and his noncounterfactual causal inference method.

4.1 Two Types of Causal Inferences

In contrast to statisticians like Donald B. Rubin who focus the analyses of causal inferences on the causal effects, Dawid proposed two problems relating causes and effects: the effects of causes and causes of effects. The effects of causes are same as the causal effects in Rubin's counterfactual model which are the comparisons between responses to different treatments. The causes of effects are defined as the treatments which cause the observed response. Here are two sentences: "*I have a headache. Will it be gone if I take aspirin?*". "*My headache has gone. Is it because I took aspirin?*". The former one is querying the effects of causes. The latter one is querying the causes of effects. Consider first the effects of causes.

4.2 Effects of Causes

4.2.1 Problems with Rubin's Counterfactual Model

In order to compare Rubin's counterfactual perspective with Dawid's noncountefactual perspective clearly, I first review some of Dawid's comments on Rubin's counterfactual model.

Jeffrey's Law

Here is one more example regarding Rubin's counterfactual model. Suppose the potential responses $(Y_1(u), Y_0(u) : u \in U)$ are modeled as independent random variables following a bivariate normal distribution with means (θ_1, θ_0), common variance ϕ_Y, and correlation $\rho \geq 0$. Then one can represent this structure by means of the mixed model:

$$Y_i(u) = \theta_i + \beta(u) + \gamma_i(u),$$

where $\beta(u)$ and $\gamma_i(u)$ are mutually independent normal random variables with means 0 and variances: $\phi_\beta = \rho\phi_Y$ and $\phi_\gamma = (1-\rho)\phi_Y$. For each u, after the treatment is applied, one of $Y_0(u)$ and $Y_1(u)$ will be the observed response, and the other will be the counterfactual. The causal effect

$$\tau(u) = Y_1(u) - Y_0(u) \sim N(\tau, 2\phi_\gamma).$$

A crucial problem with this model is that parameter ϕ_γ cannot be identified from the observed data. For instance, consider three people who have different ideas on the specific structure of this model. A thinks $\phi_\beta = 0, \phi_\gamma = \phi_Y$. B thinks $\phi_\beta = \phi_Y, \phi_\gamma = 0$. C thinks $\phi_\beta = \phi_\gamma = (1/2)\phi_Y$. Then which one is correct cannot be determined since these models are intrinsically indistinguishable on the basis of any data that could ever be observed. Only the constraint $0 \leq \phi_\gamma \leq \phi_Y$ can be obtained.

Dawid holds the point of view that any scientific investigation should follow Jeffrey's law that mathematically distinct models that cannot be distinguished on the basis of

empirical observation should lead to indistinguishable inferences. However, counterfactual models are based on unobservables: assumed observations that are not just unobserved (like missing data) but that are not even potentially observable. Thus unverifiable assumptions are intrinsic to this entire class of models including Rubin's counterfactual model. And distinct inferences based on these assumptions are drawn under indistinguishable specifications in Rubin's counterfactual model. Lord's paradox is one particular example. These violations of Jeffrey's law lead to Dawid's argument that Rubin's counterfactual model should not be used.

Fatalism

Another problem with the Rubin's counterfactual model is what Dawid terms fatalism (Dawid 2000): "*This considers the various potential responses $Y_i(u)$, when treatment i is applied to unit u, as predetermined attributes of unit u, waiting only to be uncovered by suitable experimentation.*" Rubin's counterfactual model is built on the notion that every unit has two potential responses, only one of which can be observed. And they are unaffected by any human being's activities. This idea can be (and has been) interpreted in terms of many parallel worlds (Stalnaker 1984) (Lewis 1973, 1983). Different treatments are applied in different worlds, and one can only live in one of them so that one could ever observe only one response. There is never any possibility of empirically testing this assumption of fatalism. Thus any counterfactual models including Rubin's are questionable. The commonly used STUVA assumption is a direct result of fatalism.

4.2.2 Dawid's Noncounterfactual Method

Dawid's noncounterfactual causal inference method has a fairly simple structure. First it does not assume any potential responses which may or may not be observed. That is to say it does not make use of counterfactuals. It only takes into account the actual responses in the two groups. In the context of the counterfactual model, it means

that Dawid only considers the marginal distribution of each response. As a result, there would be no unidentifiable parameters present in the analysis.

"If one cannot get a sensible answer to the question, then perhaps the question itself, with its focus on inference for $\tau(u_0)$, is not well posed." (Dawid 2000). In the analysis of causal inference, instead of asking what is the causal effect of a treatment on unit u, or what is the mean causal effect on the population of units, Dawid rephrased the question to ask which treatment would one prefer to apply to a new unit u_0. In order to answer this question, he suggests a decision-analytic approach. The idea is straightforward. Suppose there is a loss function L on the response Y. Of course, one prefers to have small $L(Y)$ on a unit. Then the conclusion is simply application of the treatment giving the smaller loss. For instance, assume $L(Y) = Y$, the decision process involves first conducting a two sample t test on the group mean responses, and then picking the treatment which gives a lower mean response if there is a significant difference. Otherwise, choose either one.

4.3 Causes of Effects

In addition to effects of causes, Dawid discussed another causal inference, causes of effects. For example, if $T = 1, Y = 1$, the causes of effects would be $P(Y_0 = 1|T = 1, Y = 1)$: the probability that the potential response of the unapplied treatment is same as the actual response. In contrast, effects of causes is the difference between the actual response and the counterfactual. It asks what are the effects the treatments cause. Dawid concluded that the counterfactual model cannot be avoided when dealing with causes of effects. As a result, unidentifiable parameters would be present in analyses.

Consider again the bivariate normal counterfacutal model in 4.2.1. The conditional distribution of $\tau(u) = Y_1(u) - Y_0(u)$, given $Y_1(u) = y$, is normal, with mean and variance

$$\lambda = E\{\tau(u)|Y_1(u) = y\} = y - \theta_0 - \rho(y - \theta_1)$$

and

$$\delta^2 = Var\left\{\tau(u)|Y_1(u) = y\right\} = (1-\rho^2)\phi_Y.$$

Only θ_1, θ_0, and ϕ_Y can be identified based on the data. The correlation between the actual responses and the counterfactuals, ρ, cannot be learned from the counterfactual model. When $\rho = 0$, $\lambda = y - \theta_0$ and $\delta^2 = \phi_Y$. If $\rho = 1$, $\lambda = \theta_1 - \theta_0$ and $\delta^2 = 0$. Assuming $\rho \geq 0$, only the constraints

$$\lambda \text{ lies between } \theta_1 - \theta_0 \text{ and } y_0 - \theta_0$$

and

$$\delta^2 \leq \phi_Y$$

can be inferred. Dawid's suggestion is to try to gain more knowledge on covariates to improve the constraints (Dawid 2000).

Chapter 5

Counterfactual vs. Noncounterfactual

While Dawid is a prominent critic of counterfactual models, there are other statisticians who voice both criticism and support for such models. Their thinking about the issues surrounding these models helps shed light on the concepts behind causal inference. In this chapter, I describe four of the more important issues.

5.1 Positivism

Positivism is a philosopical point of view which asserts that meaningful propositions must be either analytic (mathematical) or empirically falsifiable or verifiable by possible sensory observations. Dawid's comments on Rubin's counterfactual model can be viewed as based on positivism.

Some consider positivism too rigid a philosophy to serve as a suitable underpinning of scientific investigation:

- "*Logical positivism's main tenet is that meaningful propositions must be either analytic or empirically falsifiable or verifiable by possible sensory obser-*

vations.... It was too rigid and technically unworkable." (Casella, Schwartz 2000).

- "*A pragmatic empiricist insists on asking empirically testable queries, but leaves the choice of theories to convenience and imagination; the dogmatic empiricist insists on positing only theories that are expressible in empirically testable vocabulary."* (Pearl 2000).

Others, in support of Dawid, take a more positivist view. Shafer is one of them and he rephrased Dawid's perspective as : "*Dawid's central theme is that counterfactuals should be held up to de Finetti's observability criterion that it is legitimate to assess a probability distribution for a quantity Y only if Y is observable at least in principle."* (Shafer 2000).

Some argue that positivism should be abandoned in science since it is already out of fashion in philosophy: "*Starting in about 1950, logical positivism was subjected to a withering series of criticisms and has now entirely lost favor among philosophers."* (Casella, Schwartz 2000). To which Dawid replied: "*I trust, however, that my arguments will be considered on their own merits, rather than on whether they are fashionable."* (2000).

5.2 Untestable Assumptions

The other reason why Dawid rejected Rubin's counterfactual model is that inferences drawn under it are often based on untestable assumptions. However, drawing inferences based on testable assumptions is also possible in many situations under the model. For instance, under completely random assignment, the conditional expectation difference $E(Y(u)|T=1) - E(Y(u)|T=0)$ is equal to the average causal effect $E(\tau(u)) = E(Y_1(u)) - E(Y_0(u))$. Therefore, some argue that one should apply the counterfactual model in a proper way rather than simply abandon it: "*Use powerful mathematics to filter, rather than muzzle, the untestable queries that such languages*

tempt us to ask." (Pearl 2000). Dawid's reply to this is : "*My point is that the models I criticize also have untestable implications, and that it is all too easy to use them to make 'inferences' that are sensitive to purely arbitrary and untestable choices that may be made for ingredients in these models. I would prefer to build on firmer ground than this, using models that do not allow empirically meaningless statements and inferences, whenever this is possible (which currently believe is always)."* (2000).

5.3 Instrumental Use

Even though one may decide not to make causal inferences using counterfactual models, they may still have some valuable instrumental uses:

- "*The counterfactual modeling languages are somewhat richer than the ones needed for routine predictions. . . . lack of identifiability is possible in any statistical problem. But it should not prevent us from using counterfactuals to provide simple and clear explanations for causality."* (Pearl 2000).
- "*The meaningfulness of counterfactual variables need not prevent one from using them for mathematical convenience."* (Shafer 2000).

But Dawid did not agree with this: "*. . . However, I do not feel that the counterfactual approach to causal inference has, as yet, provided any of these advantages. "* (2000).

5.4 Applications

Dawid discussed his noncounterfactual approach only in the context of controlled experiments. In contrast, the counterfactual model has shown its great power of clarifying and solving problems in a variety of settings and areas. Lord's paradox is a good example. Although Dawid said that he is currently developing the ideas, it

seems that his noncounterfactual model cannot affect Rubin's counterfactual model's dominance until some ideas are well established under it.

Chapter 6

Conclusion

In this report, I mainly introduce two currently competing views on causal inference: the counterfactual perspective and noncounterfactual perspective. It is shown that one of the most popular counterfactual models, Rubin's counterfactual model, has great usefulness. As a result, although even Rubin and his advocates do not dispute the legitimacy of Dawid's concerns on the counterfactual perspective, Rubin's model is still widely used and seems to have a dominant position among all the models regarding causal inference. In contrast, Dawid's noncounterfactual causal inference method is no doubt based on a firmer ground. But its application is highly limited and still needs much development.

Bibliography

[1] Cox, D. R. (1958). The Planning of Experiments, New York: John Wiley.

[2] Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95, 407-424.

[3] Fisher, R. A. (1926). The Arrangement of Field Experiments. *Journal of Ministry of Agriculture*, 33, 503-513.

[4] Holland, P. W. and Rubin, D. B. (1983). On Lord's paradox. In H. Wainer and S. Messick (Eds.), *Principals of modern [End Page 352] psychological measurement: A festschrift for Frederic M. Lord (pp. 3-26)*. Hillsdale, NJ: Erlbaum.

[5] Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*,81,945-970.

[6] Hume, D. (1740). A Treatise on Human Nature.

[7] Hume, D. (1748). An Inquiry Concerning Human Understanding.

[8] Kempthorne, 0. (1952). The Design and Analysis of Experiments, New York: John Wiley.

[9] Lewis, D. (1973). Counterfactuals, Cambridge, MA: Harvard University Press.

[10] Locke, J. (1990). An Essay concerning Human Understanding, Book II, Chapter XXVI.

[11] Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 68, 304-305.

[12] Mill, J. S. (1843). A System of Logic.

[13] Neyman, J. (with Iwaszkiewicz, K., and Kolodziejczyk, S.) (1935). *Statistical Problems in Agricultural Experimentation (with discussion), Supplement of Journal of the Royal Statistical Society*, 2, 107-180.

[14] Stalnaker, R. (1984). Inquiry, Boston, MA: Bradford Books.

[15] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66,688-701.

[16] Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 7,34-58.

[17] Rubin, D. B. (2004). Teaching Statistical Inference for Causal Effects in Experiments and Observational Studies. *Journal of Educational and Behavioral Statistics*, 3,343-367.

[18] Rubin, D. B. (1990). [On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.] Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, Vol. 5, No. 4, 472-480.