UNIVERSAL FORMULAS FOR TREATMENT EFFECTS FROM NONCOMPLIANCE DATA

Alexander A. Balke and Judea Pearl

University of California at Los Angeles

This paper establishes formulas that can be used to bound the actual treatment effect in any experimental study in which treatment assignment is random but subject compliance is imperfect. These formulas provide the tightest bounds on the average treatment effect that can be inferred given the distribution of assignments, treatments, and responses. Our results reveal that even with high rates of noncompliance, experimental data can yield significant and sometimes accurate information on the effect of a treatment on the population.

1 Introduction

Consider an experimental study where random assignment has taken place but compliance is not perfect (i.e., the treatment received differs from that assigned). It is well known that under such conditions a bias may be introduced, in the sense that the true causal effect of the treatment may deviate substantially from the causal effect computed by simply comparing subjects receiving the treatment with those not receiving the treatment. Because the subjects who did not comply with the assignment may be precisely those who would have responded adversely (positively) to the treatment, the actual effect of the treatment, when applied uniformly to the population, might be substantially less (more) effective than the study reveals.

In an attempt to avert this bias, analysts sometimes resort to parametric models which make restrictive commitments to a particular mode of interaction between compliance and response [Efron and Feldman, 1991]. [Angrist *et al.*, 1993] have identified a set of assumptions under which a nonparametric correction formula, called "Instrumental Variables", is valid for certain subpopulations. These subpopulations cannot be identified from empirical observation alone, and the need remains to devise alternative, assumption-free formulas for assessing the effect of treatment over the population as a whole. [Robins, 1989] and [Manski, 1990] derived nonparametric bounds on treatment effects which are assumption-free, but do not make full use of the information available in the data. In this paper, we provide the tightest possible bounds on the average treatment effect. These bounds rely solely on observed quantities and are universal, that is, valid no matter what model actually governs the interactions between compliance and response.

2 Problem formulation

The canonical partial-compliance setting can be graphically modeled as shown in Figure 1.

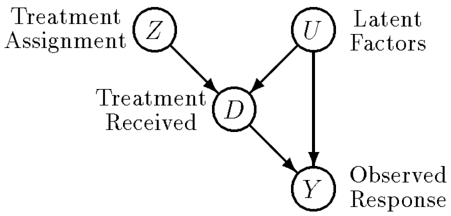

Figure 1: *Graphical representation of causal dependencies in a randomized clinical trial with partial compliance.*

We assume that Z, D, and Y are observed binary variables where Z represents the (randomized) treatment assignment, D is the treatment actually received, and Y is the observed response. U represents specific characteristics of an individual subject, namely, all factors, both observed and unobserved, that may influence a subject's choice of treatment D and eventual outcome Y. To facilitate the notation, we let z, d, and y represent, respectively, the values taken by the variables Z, D, and Y, with the following interpretation: $z \in \{z_0, z_1\}$, z_1 asserts that treatment has been assigned (z_0, its negation); $d \in \{d_0, d_1\}$, d_1 asserts that treatment has been administered (d_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 model analyzed invokes two assumptions of independence:

1. For a given individual, the treatment assignment does not influence Y directly, but only through the actual treatment D, that is,

$$Z \perp\!\!\!\perp Y \mid \{D, U\} \tag{1}$$

   In practice, any direct effect Z might have on Y would be adjusted for through the use of a placebo.

2. Z and U are marginally independent, that is, $Z \perp\!\!\!\perp U$. This independence is partly ensured through the randomization of Z, which rules out a common cause for both Z and U. Additionally, $Z \perp\!\!\!\perp U$ implies that the assignment Z does not alter a person's initial characteristics U. This still permits the final choice of treatment D to be dependent on the subject's initial reaction to the treatment assigned Z, as long as the initial reaction can be regarded as transitory, i.e., not having direct effect on Y.

These assumptions impose on the joint distribution[1] the decomposition

$$P(y, d, z, u) \;=\; P(y|d, u)\; P(d|z, u)\; P(z)\; P(u) \tag{2}$$

which, of course, cannot be observed directly because U is a latent variable. However, the marginal distribution $P(y, d, z)$ and, in particular, the conditional distributions $P(y, d|z)$, $z \in \{z_0, z_1\}$, are observed, and the challenge is to assess the causal effect of D on Y from these distributions.[2]

[1] We take the liberty of denoting the prior distribution of U by $P(u)$, even though U may consist of continuous variables. If U is sufficiently refined, the term $P(y|d, u)$ will become deterministic, but such refinement is not necessary for our analysis.

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

Causal analysis of treatment effects formalizes changes in the joint distribution which are induced by external interventions [Pearl, 1994]. In particular, the analysis concerns the effect of *local* interventions, relative to which $P(y|d,u)$ is a stable quantity, i.e., the probability that an individual with characteristics $U = u$ given treatment $D = d$ will respond with $Y = y$ remains the same, regardless of how the treatment was selected — be it by choice, persuasion, or policy. Thus, to predict the distribution of Y if the treatment D were applied uniformly to the population, we should calculate

$$P(y|D = d \text{ uniformly}) \stackrel{\triangle}{=} P(y^*|\hat{d}^*) \stackrel{\triangle}{=} \sum_u P(y|d,u)P(u) \tag{3}$$

where $P(y^*|\hat{d}^*)$ stands for the probability that Y would have been equal to y, if D were (counterfactually) equal to d. A value annotated with a superscript asterisk $(*)$ denotes the value of the corresponding variable in a hypothetical counterfactual world, and a value annotated with a hat $(\hat{\ })$ indicates that the corresponding variable has been set to the value by an external local action. Details of the semantics and evaluation of counterfactual probabilities may be found in [Balke and Pearl, 1994].

If we are interested in estimating the average *change* in Y due to treatment, we can similarly define the average *causal effect*, $\text{ACE}(D \rightarrow Y)$ ([Holland, 1988]), as

$$\text{ACE}(D \rightarrow Y) \quad = \quad P(y_1^*|\hat{d}_1^*) - P(y_1^*|\hat{d}_0^*) \tag{4}$$

The task of causal inference is then to estimate or bound the expressions in Eqs. (3) and (4), given the observed probabilities $P(y,d|z_0)$ and $P(y,d|z_1)$. This may be accomplished by a procedure detailed in [Balke and Pearl, 1994], which is based on linear programming optimization coupled with the fact that the domain of U can be partitioned into sixteen equivalent classes, each representing one of four possible mappings from Z to D conjoined with one of four possible mappings from D to Y.

3 Results

Let the conditional distribution $P(y,d|z)$ over the observable variables be notated as follows:

$$\begin{array}{ll}
p_{00.0} = P(y_0, d_0|z_0) & p_{00.1} = P(y_0, d_0|z_1) \\
p_{01.0} = P(y_0, d_1|z_0) & p_{01.1} = P(y_0, d_1|z_1) \\
p_{10.0} = P(y_1, d_0|z_0) & p_{10.1} = P(y_1, d_0|z_1) \\
p_{11.0} = P(y_1, d_1|z_0) & p_{11.1} = P(y_1, d_1|z_1)
\end{array}$$

Symbolic optimization of Eq. (3) leads to the following bounds:

$$\max\left\{\begin{array}{c} p_{10.1} \\ p_{10.0} \\ p_{10.0}+p_{11.0}-p_{00.1}-p_{11.1} \\ p_{01.0}+p_{10.0}-p_{00.1}-p_{01.1} \end{array}\right\} \leq P(y_1^*|\hat{d}_0^*) \leq \min\left\{\begin{array}{c} 1-p_{00.1} \\ 1-p_{00.0} \\ p_{01.0}+p_{10.0}+p_{10.1}+p_{11.1} \\ p_{10.0}+p_{11.0}+p_{01.1}+p_{10.1} \end{array}\right\}$$

and

$$\max\left\{\begin{array}{c} p_{11.0} \\ p_{11.1} \\ -p_{00.0}-p_{01.0}+p_{00.1}+p_{11.1} \\ -p_{01.0}-p_{10.0}+p_{10.1}+p_{11.1} \end{array}\right\} \leq P(y_1^*|\hat{d}_1^*) \leq \min\left\{\begin{array}{c} 1-p_{01.1} \\ 1-p_{01.0} \\ p_{00.0}+p_{11.0}+p_{10.1}+p_{11.1} \\ p_{10.0}+p_{11.0}+p_{00.1}+p_{11.1} \end{array}\right\}$$

In addition, one can prove that these are the tightest possible assumption-free bounds and, moreover, the tightest bounds on the average treatment effect are obtained by taking differences of the bounds corresponding to the individual terms in Eq. (4). The bounds shown above constitue substantial improvement over those derived by [Robins, 1989] and [Manski, 1990], which correspond to the two upper terms in each expression. The width of these bounds cannot exceed the rate of noncompliance, $P(d_1|z_0) + P(d_0|z_1)$, and may in some cases collapse to a point estimate, even when as many as 50% of subjects switch over to unassigned treatments.

Examples and additional results regarding bounds on treatment effects in partial compliance studies are presented in [Balke and Pearl, 1993]. In particular, it is shown that the two independence assumptions underlying randomized-assignment experiments, although not directly testable, impose the following testable constraints on the observed distribution:

$$\begin{aligned}
P(y_0, d_0|z_0) + P(y_1, d_0|z_1) &\leq 1 \\
P(y_0, d_1|z_0) + P(y_1, d_1|z_1) &\leq 1 \\
P(y_1, d_0|z_0) + P(y_0, d_0|z_1) &\leq 1 \\
P(y_1, d_1|z_0) + P(y_0, d_1|z_1) &\leq 1
\end{aligned}$$

If any of these inequalities is violated, the investigator can safely deduce that either the assignments were not properly randomized, or treatment assignment exerts some direct influence on subjects' responses.

Finally, the method of causal analysis outlined above permits us to evaluate a variety of more intricate counterfactual probabilities. For example, suppose we wish to assess the probability that a given individual would have recovered had he/she not been assigned treatment (z_0), when in actuality he/she has been assigned the treatment (z_1), taken the treatment (d_1), and not recovered (y_0). In this case, we analyze the causal effect of the assignment in the subpopulation characterized by $\{z_1, d_1, y_0\}$, which leads to the following bounds:.

$$\frac{1}{p_{01.1}} \max \left\{ \begin{array}{c} 0 \\ p_{01.1} - p_{00.0} - p_{01.0} \\ p_{10.0} - p_{10.1} - p_{11.1} \end{array} \right\} \leq P(y_1^* | \hat{z}_0^*, z_1, d_1, y_0) \leq \frac{1}{p_{01.1}} \min \left\{ \begin{array}{c} p_{10.0} \\ p_{01.1} \\ 1 - p_{01.0} - p_{11.1} \\ 1 - p_{00.0} - p_{10.1} \end{array} \right\}$$

A general method for obtaining such bounds is detailed in [Balke and Pearl, 1994].

4 Conclusion

In an attempt to avert confounding bias in randomized studies involving noncompliance, analysts usually advocate the use of "intent-to-treat" analysis, which compares assignment groups regardless of the treatment actually received. Estimates based on such analysis are free of confounding bias 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, treatment effectiveness is determined by the average causal effect, $E_u[P(y_1|u, d_1) - P(y_1|u, d_0)]$ for which we have provided universal and strict bounds. The formulas established in this paper should enable the analyst to determine the extent to which estimates based on intent-to-treat analysis may deviate from the actual treatment effect. This information should be useful for assessing whether efforts to ensure population compliance have the potential of increasing the overall benefit of the treatment.

Acknowledgements

The research was partially supported by Air Force grant #AFOSR 90 0136, NSF grant #IRI-9200918, Northrop Micro grant #92-123, and Rockwell Micro grant #92-122. Alexander Balke was supported by the Fannie and John Hertz Foundation.

References

[Angrist *et al.*, 1993] J.D. Angrist, G.W. Imbens, and D.B. Rubin. Identification of causal effects using instrumental variables. Technical Report No. 136, Department of Economics, Harvard University, Cambridge, MA, June 1993.

[Balke and Pearl, 1993] Alexander Balke and Judea Pearl. Nonparametric bounds on causal effects from partial compliance data. Technical Report R-199, Cognitive Systems Laboratory, Computer Science Department, UCLA, September 1993. Submitted.

[Balke and Pearl, 1994] Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA*, pages 46–54, San Fransisco, CA, 1994. Morgan Kauffman.

[Efron and Feldman, 1991] B. Efron and D. Feldman. Compliance as an explanatory variable in clinical trials. *Journal of the American Statistical Association*, 86(413):9–26, March 1991.

[Holland, 1988] Paul W. Holland. Causal inference, path analysis, and recursive structural equations models. In C. Clogg, editor, *Sociological Methodology*, pages 449–484. American Sociological Association, Washington, DC, 1988.

[Manski, 1990] Charles F. Manski. Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80:319–323, May 1990.

[Pearl, 1994] Judea Pearl. Causal diagrams for empirical research. Technical Report R-218-L, Revision I, UCLA Cognitive Systems Laboratory, May 1994. To appear in *Biometrika*.

[Robins, 1989] J.M. Robins. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In L. Sechrest, H. Freeman, and A. Mulley, editors, *Health Service Research Methodology: A Focus on AIDS*, pages 113–159. NCHSR, U.S. Public Health Service, 1989.

Cognitive Systems Laboratory
6291 Boelter Hall
University of California
Los Angeles, CA 90024

E-mail:
<balke@cs.ucla.edu>
<judea@cs.ucla.edu>