COVARIATE SELECTION: A SIMPLE SOLUTION TO A LONG-STANDING PROBLEM

Judea Pearl, University of California, Los Angeles
Computer Science Department, 4532 Boelter Hall, Los Angeles, CA 90024-1596

Key Words: Causality, Simpson's paradox, confounding

Whenever we undertake to evaluate the effect of one factor (X) on another (Y), the question arises as to whether we should adjust our measurements for possible variations in some other variable, Z, sometimes called "covariate", "concomitant" or "confounder". Adjustment amounts to partitioning the population into groups that are homogeneous relative to Z, assessing the effect of X on Y in each homogeneous group and, finally, averaging the results. The importance of such adjustment has been recognized as early as 1899, when Karl Pearson, the founder of modern statistics, discovered what in modern terms would be called the Simpson's paradox: Any statistical relationship between two variables may be reversed or negated by including additional factors in the analysis.

The classical case demonstrating Simpson's reversal is the study of Berkeley's alleged sex bias in graduate admission [Bickel *et al.*, 1975], where overall, data show a higher rate of admission among male applicants but, broken down by departments, data show a slight bias toward female applicants.

Despite a century of analysis, the Simpson's reversal phenomenon continues to "trap the unwary" [Dawid, 1979] and the main question – whether an adjustment for a given covariate Z is appropriate in any given study – continues to be debated informally, on a cases by case basis, resting on folklore and intuition rather than on hard mathematics. The statistical literature is remarkably silent on this issue and, aside from the standard advice that one should not adjust for a covariate that is affected by the putative cause (X), it provides no further guidelines as to what covariates would be admissible for adjustment and what assumptions would be needed for making this this determination.[1] In the Berkeley story, for example, if we suspect that a certain undisclosed attribute of a candidate (e.g., music aptitude) has influenced both the admission decision and applicants' choice of departments, then it is not quite clear whether adjusting for department choice would reduce or increase bias in assessing alleged sex discrimination in admissions. More significantly, it is not uncommon to find trained statisticians disagreeing on such decisions, even if the confounding attribute can safely be assumed gender-independent. A concrete controversy that arose out of the problem of covariate adjustment is the episode of "reverse regression" which occupied the social science literature in the 1970's: Should we, in salary discrimination cases, compare salaries of equally qualified men and women or, instead, compare qualifications of equally paid men and women. Another controversy, known as the Lord Paradox, originated with the question of whether we should adjust for socioeconomic status in assessing the effect of tutoring program on reading score, knowing that children from well to do families, who do better on the reading test, are more likely to be in the program Lord paradox, in its various manifestations, still lingers on, in the psychometric literature [Wainer, 1991].

The primary reason for the persistence of confusion over the appropriateness of statistical adjustment is that the answers depend on causal assumptions, and statisticians, by and large, are reluctant to discuss such assumptions forthrightly. First, causal assumptions (e.g., that the rooster's call does not cause the sun to rise) are perceived less objective than typical statistical assumptions (e.g., that two variables are correlated), presumably because the former cannot be established by observational studies. Second, even when causal facts are established to the satisfaction of statistical standards (say, by subjecting our rooster to controlled randomized experiment), we still lack an adequate mathematical notation to state those assumptions symbolically – causal relationships cannot be distinguished from statistical associations in the standard language of probability theory.

The purpose of this paper is to report a formal and general solution to the problem of statistical adjustment using the language of graphs. This language [Pearl, 1995], which can be traced back to the

[1] Most of the statistical literature is satisfied with informal warnings that "Z should be quite unaffected by X" [Cox, 1958, page 48], which is necessary but not sufficient, or that X should not precede Z [Shafer, 1995, page 294], which is neither necessary nor sufficient. In some academic circles, a criterion called "ignorability" is invoked [Rosenbaum and Rubin, 1983], which merely paraphrases the problem in the language of counterfactuals. Simplified, it reads: Z is an admissible covariate relative to the effect of X on Y if, for every x, the value that Y would obtain had X been x is conditionally independent of X, given Z. Since counterfactuals are not observable, and judgments about conditional independence of counterfactuals are not readily assertable from ordinary understanding of causal processes, ignorability has remained a theoretical construct, with only minor impact on practice. Practicing epidemiologists, for example are still debating the meaning of "confounding", and often adjust for wrong sets of covariates [Weinberg, 1993].

geneticist Sewal Wright (1919), permits the investigator to express causal assumptions in the form of arrows among quantities of interest, and, once the graphs are completed, a simple procedure would decide whether a proposed adjustment is appropriate relative to the quantity under evaluation.

The procedure is described in the following five steps, which determine whether a set of variables Z should be adjusted for, when we we wish to evaluate the total effect of X on Y.

Procedure:[2]

Input: Directed acyclic graph in which three subsets of nodes are marked X, Y and Z.

Output: A decision whether the effect of X on Y can be determined by adjusting for Z.

Step 1. Exit with failure if any node in Z is a descendant of X,

Step 2. (simplification) Simplify the diagram by eliminating all nodes (and their incident edges) which are not ancestors of either X, Y or Z.

Step 3. (triangulation) Add an undirected edge between any two ancestors of Z which share a common child.

Step 4. (pruning) Eliminate all arrows emanating from X.

Step 5. (symmetrization) Strip the arrows from all directed edges.

Step 6. (test) If, in the resulting undirected graph, Z intercepts all paths between X and Y, then Z is an appropriate covariate for statistical adjustment. Else, Z should not be adjusted for.

When failure occurs in Step 1, it does not mean that the measurement of Z cannot be useful for estimating the effect of X on Y; nonstandard adjustments might then be used instead of the standard method of partitioning into groups homogeneous relative to Z (see [Galles and Pearl, 1995]). Finally, if the objective of the study is to evaluate the "direct", rather than the "total" effect of X on Y, as is the case in the Berkeley example, then other graphical procedures are available to determine the appropriate adjustment (see [Pearl and Robins, 1995]), which embody the requirement that other factors (of Y) should be "held constant".

[2] This procedure is an adaptation of the back-door criterion in [Pearl, 1995], using the triangulation test [Lauritzen *et al.*, 1990] of d-separation [Pearl, 1988].

References

[Bickel *et al.*, 1975] P.J. Bickel, E.A. Hammel, and J.W. O'Connell. Sex bias in graduate admissions: Data from berkeley. *Science*, 187:398–404, 1975.

[Cox, 1958] D.R. Cox. *The Planning of Experiments.* John Wiley and Sons, NY, 1958.

[Dawid, 1979] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series A*, 41:1–31, 1979.

[Galles and Pearl, 1995] D. Galles and J. Pearl. Testing identifiability of causal effects. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 185–195. Morgan Kaufmann, San Francisco, CA, 1995.

[Lauritzen *et al.*, 1990] S.L. Lauritzen, B.N. Dawid, A.P. abd Larsen, and H.G. Leimer. Independence properties of directed markov fields. *Networks*, 20(5):491–505, 1990.

[Pearl and Robins, 1995] J. Pearl and J.M. Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 444–453. Morgan Kaufmann, San Francisco, CA, 1995.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligence Systems.* Morgan Kaufmann, San Mateo, CA, 1988. (Revised 2nd printing, 1992).

[Pearl, 1995] J. Pearl. Causal diagrams for experimental research. Technical Report R-218-B, Computer Science Department, UCLA, 1995. To appear in *Biometrika*, December 1995.

[Rosenbaum and Rubin, 1983] P. Rosenbaum and D. Rubin. The central role of propensity score in observational studies for causal effects. *Biometrica*, 70:41–55, 1983.

[Shafer, 1995] G. Shafer. *The Art of Causal Conjecture*. MIT Press, Cambridge, MA, 1995. Forthcoming.

[Wainer, 1991] H. Wainer. Adjusting for differential base-rates: Lord's paradox again. *Psychological Bulletin*, 109:147–151, 1991.

[Weinberg, 1993] C.R. Weinberg. Toward a clearer definition of confounding. *American Journal of Epidemiology*, 137(1):1–8, 1993.