1 The Intuition Behind Inverse Probability Weighting

Michael Foster writes:

I'm an economist here in the UNC school of public health and trying to work on the intuition of MSM for my non-methodologists collaborators. My bios and epi colleagues can give me mechanical answers but are short on intuition at times. Here are two questions:

1. Consider a regressor that is a confounding variable but that is also a victim of unobserved confounding itself. Why does weighting with this troublesome covariate not cause bias that regression causes (collider bias)? In this case, I'm principally thinking about past exposures and how to handle them in an analysis of dynamic treatment. Marginal structural models (MSM) including them in calculating the weights; Robins suggests that including them as covariates in the outcome equation produces the "null paradox."

   Here's my answer. A confounding variable has two characteristics–it is related to the exposure and to the outcome. When we weight with that variable, we break the link between the exposure and that variable. However, other than the portion due to the exposure, we do not eliminate the relationship between the covariate and the outcome. In that way (by not breaking both links), we avoid the bias created by the collider issue.

2. How do I know what variables to include in the numerator of the MSM weight?

   Here's my answer: I would include in the weights those variables that will be included in the analysis of the outcome. Their presence in the denominator of the weight is essentially duplicative–we're accounting for them there and in the outcome model.

Comment, by Judea Pearl

Your question deals with the intuition behind "Inverse Probability Weighting" (IPW), an estimation technique used in several frameworks, among them Marginal Structural Models (MSM). However, the division by the propensity score $P(X = 1|Z = z)$ or the probability of treatment $X = 1$ given observed covariates $Z = z$, is more than a step taken to improve estimation; it is dictated by the very definition of "causal effect," and appears therefore, in various guises, in every method of effect estimation – it is a property of Nature, not of our efforts to unveil the secrets of Nature.

Let us first see how this probability ends up in the denominator of the effect estimand, and then deal with the specifics of your question, dynamic treatment and unobserved confounders.

Suppose we had a treated subject with propensity score of 0.2, then for every such treated subjects there would be 4 control subjects. If we want to consider what would happen if everyone received treatment then we need to consider what happens to the other four control subjects once they are assigned treatment. How many of them will have outcome $Y = 1$ (say recovery) and how many outcome $Y = 0$.

The best way to answer this question is to decompose the distribution of subjects into three sub-distributions, each representing a process the determines the probability of

another variable. Specifically, the pre-intervention distribution $P(Y = 1, X = 1, Z = z)$ can be decomposed into a product of three conditional probabilities:

$$P(Y = 1, X = 1, Z = z) = P(Y = 1|X = 1, Z = z) * P(X = 1|Z = z) * P(Z = z)$$

where

$P(Y = 1|X = 1, Z = z)$ describes how the outcome depends on treatment and covariate Z.

$P(X = 1|Z = z)$ describes how subjects choose treatment prior to intervention.

$P(Z = z)$ describes the prior distribution of the covariates.

Call the post-intervention distribution P', and decompose it into the same three mechanisms,

$$P'(Y = 1, X = 1, Z = z) = P'(Y = 1|X = 1, Z = z) * P'(X = 1|Z = z) * P'(Z = z)$$

we ask which of the three remains invariant. Clearly, the last mechanism is invariant $P'(Z = z) = P(Z = z)$ because Z is a pre-treatment covariate. Likewise, the outcome mechanism is invariant

$$P'(Y = 1|X = 1, Z = z) = P(Y = 1|X = 1, Z = z)$$

because Z is assumed to be the complete set of confounders, so, Y depends on X and Z plus random noise that is not affected by the treatment. Finally, the treatment mechanism itself undergoes a drastic change, with $P'(X = 1|Z = z) = 1$ because everyone gets the treatment

As a result, we see that the post-intervention distribution is equal to the product:

$$P'(Y = 1, X = 1, Z = z) = P(Y = 1|X = 1, Z = z) * P(Z = z) \tag{1}$$

or the function

$$P'(Y = 1, X = 1, Z = z) = \frac{P(Y = 1, X = 1, Z = z)}{P(X = 1|Z = z)} \tag{2}$$

At this point, two perspectives offer themselves for analysis: we can either view the post-treatment distribution P' as a mutilated version of the pre-treatment distribution P, with the treatment-probability term excised, or we can view each cell (or subject) characterized by the tuple $(Y = 1, X = 1, Z = z)$ to be assigned a new weight, given by the old weight times the inverse probability of the treatment. The first perspective (Eq. (1)) leads to estimation by regression, stratification, or propensity score (PS) matching, while the second (Eq. (2)) leads to Inverse Probability Weighing (IPW). The asymptotic equivalence of the two approaches is assured by the equality of (1) and (2) and, therefore, questions that pertain to asymptotic analysis, for example, whether a bias is removed, whether a given covariate should be "weighted on" etc, should be addressed not to the estimation technique but to the assumptions embodied in the model; the relation between P and P' is dictated by those assumptions and those assumptions only.

This partially answers your first question:

Why does weighting with this troublesome covariate not cause bias that regression causes (collider bias)?

Answer: Weighting with any covariate will ALWAYS cause the same bias that regression causes. "Bias" is an asymptotic concept, hence it is invariant to shifting from regression to weighting. "Collider bias," for example, cannot be eliminated by either method. (BTW, a confounder need NOT be a covariate "related to both exposure and outcome," nor need such a covariate be a confounder (*Causality*, p. 195). These definitions are outdated.)

To complete this part of the discussion, we should also note that the inverse probability weighting (Eq. 2) can be given a counting interpretation. Suppose there are N_z subjects in the $Z = z$ stratum. Among these N_z subjects, those who chose treatment and recovered count to $N_z P(Y = 1, X = 1|Z = z)$ subjects. The number of such subjects in the post-intervention distribution is:

$$N_z P'(Y = 1, X = 1|Z = z) = N_z \frac{P(Y = 1, X = 1|Z = z)}{P(X = 1|Z = z)}$$

We see that, while in the pre-intervention data, there are only $N_z P(Y = 1, X = 1|Z = z)$ subjects in stratum $Z = z$ who take treatment and recover, their number would increase by the inverse probability factor

$$1/P(X = 1|Z = z)$$

if everyone is compelled to undergo treatment. Thus for example, if $P(X = 1|Z = z) = 0.2$, then every subject in our dataset represents five subjects in the post-interventional data, with identical treatment-outcome behavior.

This completes the first part of my comments which demystifies, I hope, the appearance of the probability term in the denominator of causal effect estimators. It is based on the modern conception of causal effects in terms of mechanism invariance and mechanism disabling, also called "surgery" (see *Causality* (Pearl, 2009, Chapter 3), or my survey paper http://ftp.cs.ucla.edu/∼kaoru/r355-tr.pdf) Other explanations are floating around, some based on conditional independence of counterfactuals, some on "ignorability," "comparability," "heterogeneity," "exogeneity," "exchangeability," and perhaps more...; they all boil down to surgery but lack its generated and intuitive appeal.

I believe the confusion that you have noticed among you collaborators may be attributed to the slow penetration of modern concepts of causation into mainstream literature and education. The purpose of this blog is to speed up the process.

2 Dynamic treatment and multiple confounders

The principles of surgery and invariance permits us to generalize the computation of the post-intervention distribution to any system of multiple treatments and multiple covariates, some affecting treatments and some affected by treatments, as long as we assume that all confounders are observable. The idea is always to decompose the pre-intervention distribution into a product of conditional probabilities representing causal mechanisms and remove from the product those factors that correspond to specific interventions. The

resulting expression is called "truncated product" in *Causality* 2009, and also goes by the names Manipulation Theorem (Spirtes et al., 2000), G-computation (Robins, 1986), and perhaps others.

To demonstrate, assume that we have two covariates, Z_1 and Z_2, two treatments, X_1 and X_2 and one outcome, Y. Assume further that the temporal order of these variables is (Z_1, X_1, Z_2, X_2, Y) and that every variable is affected by all its predecessors. Suppose that we are interested in the effect of setting treatments X_1 and X_2 to levels x_1 and x_2 respectively. To compute the post-intervention distribution $P'(Z_1, Z_2, Y)$ we need not guess what goes into the denominator or numerator; it all follows from the truncated product expression.

$$P'(Z_1, X_1 = x_1, Z_2, X_2 = x_2, Y) = \frac{P(Z_1, X_1 = x_1, Z_2, X_2 = x_2, Y)}{P(X_1 = x_1|Z_1)P(X_2 = x_2|Z_1, X_1 = x_1, Z_2)}$$

The denominator is simply the product of all conditional probability terms that correspond to the local treatment variables, in our case X_1 and X_2. The fact that Z_2 is affected by the treatment $X_1 = x_1$ need not concern us. We need not be inhibited by cautionary guidelines that were devised by statisticians in the pre-causal era (e.g., not to control for covariates that are affected by the treatment (Cox, 1958, p. 58)) These were useful guidelines in their times, and have since been replaced by modern rules, formulas and understanding.

3 Extension to unobserved confounders.

If some of the confounders are unobserved, the causal effect of interest may not be identifiable, and then no procedure in the world can estimate it consistently, not even those which are wrongly believed to do so (e.g., propensity score matching, or MSM, or others). Today we have a complete procedure for determining whether a partially observed set of covariates is sufficient to identify a given causal effect. (See *Causality* (Pearl, 2009, p. 105)).

Thus, to re-address your first question about unobserved confounding:

> "Consider a regressor that is a confounding variable but that is also a victim of unobserved confounding itself,"

we would need to see the graph before answering this question, because the answer depends critically on *how* the "victim" is "victimized."

The mixed motivation (estimation and identification) through which IPW has been introduced to causal analysis has created a myth that weighing can simulate physical interventions and thus solve all problems with unobserved confounders. In your words:

> "When we weight with that variable, we break the link between the exposure and that variable."

This is not the case. While certain interventions can, under certain assumptions be simulated by weighing, not every weighing simulates an intervention, especially not under conditions of unmeasured confounders. The safe way to practice IPW is to start with the

desired intervention, simulated it through the truncated product decomposition (as in Eq. (1)) and examine what weighting scheme (if any) is implied by it.

There is no simple weighting formula for handling a general problem with unobserved confounders, though there is a simple formula that computes P' from P whenever it is feasible, and a simple rule that detects when it is infeasible. Show us the graph, and you will get the answer in five seconds (assuming I am awake).

Regarding your second question:
"How do I know what variables to include in the numerator of the MSM weight?" Just follow the truncated product rule – you will never go astray, guarantee!!!

References

Cox, D. (1958). *The Planning of Experiments*. John Wiley and Sons, NY.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling* **7** 1393–1512.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search*. 2nd ed. MIT Press, Cambridge, MA.