Algebraic Ground Truth Inference for Counterfactual Assumptions

Andrés Corrada-Emmanuel
Swoop
andres@swoop.com

Aditya Chaganti
Swoop
aditya@swoop.com

Edward Pantridge
Swoop
ed@swoop.com

Edward Zahrebelski
Swoop
eddie@swoop.com

Simeon Simeonov
Swoop
sim@swoop.com

Abstract

Scientific computations in Causal Inference are no different than measurements using physical scientific instruments. They have uncertainty and their quality needs benchmarking to prevent misuse in their intended context. We discuss a non-parametric methodology that can be used to check the validity of counterfactual assumptions in a recent paper by Cinelli and Pearl [1]. We find that their assumptions of equality between counterfactual quantities across three domains can be relaxed and their own data confirms their validity to better than one part in 10000. The general ideas of the method, algebraic ground truth inference, are discussed to encourage the development of other calculations that could, for example, verify how close to independence U errors are in a particular context in a model independent fashion.

1 Introduction

IPM.ai is a US start-up working in the domain of medical AI. Causal Inference (CI) offers a systematic and principled methodology for us to handle a variety of quality and application issues we face. Some of these are - correcting for sample biased data, and transferring conclusions across domains (Bareinboim and Pearl [2] and references therein). CI can also be used to improve medical ML models by incorporating counterfactual features, Richens et al. [3].

However, with CI we face the same problems as any other scientific theory that seeks to explain and generalize observed natural phenomena. How are we to test theories absent meta-theories about them? We face the same problems faced by Regnault in the 18th century when trying to establish the validity of different instruments for measuring temperature (Chang [4]). How can we trust the output of an instrument without invoking another theory about how the instrument works? His solution was to invoke a radical empiricism that rejected all theory and instead relied on the study of the output of an ensemble of instruments. We provide a modern version of this idea that is based on algebra and illuminates the advantages and limitations of such a non-parametric approach. We use it to solve what we readily will accept as a "toy experiment" - can data be used to test the counterfactual equality assumptions used by Cinelli and Pearl [1]? Our answer will be yes and the final result is shown in Table 1. These results are satisfying given our knowledge of the relation between these three domains. They basically state that we are more justified in believing that Aceh and West Java have similar

Preprint. Under review.

pair	Δ_{01}
$\Delta_{01,\text{Aceh,West Java}} = 0$	0.000053
$\Delta_{01,\text{Aceh,Sarlahi}} = 0$	0.00013
$\Delta_{01,\text{Sarlahi,West Java}} = 0$	0.000088

Table 1: Deviations of PS_{01} counterfactual assumptions from equality assuming all PS_{10} counterfactuals are equal

counterfactuals than either one is equal to the one for Sarlahi. We will explain how this calculation came about in what follows.

Our interest in this matter is related to a similar problem in the problem of self-assessing binary classifiers or regressors when we do not have the ground truth of their correct answers Corrada-Emmanuel et al. [5]. Scientific models must make simplifying assumptions and we have an intuitive sense that most unknown variables that could affect a given phenomena under study are irrelevant. This way this appears in Causal Inference is as the assumption that the variance in observational data due to unknown U variables is independent among them. It is unsatisfying to believe that one, we must assume this independence and cannot confirm it experimentally and two, we cannot deal with what is expected to be the practical norm - small, nearly uncorrelated errors due to the U variables.

1.1 Relation to previous work

It may seem strange that we could deduce properties of unknown quantities for which we have no theory or knowledge. But it has been done before - the Good-Turing estimate for the number of unseen species. Similarly, there is an algebraic, non-parametric methodology that can measure the performance of classifiers and regressors on a sample without ground truth knowledge of what their correct answers should have been.

There are three topics familiar to the CI community that are similar to the approach advocated here but have different goals and approaches - identification of causal quantities, sensitivity analysis, and algebraic statistics. We discus each in turn for its similarities and differences to algebraic GTI.

Since causal models may not be fully known in a particular setting with given data, the problem of *identifiability of causal quantities* is to ascertain what causal quantities can be computed (bounds or point estimates) given partial knowledge of the model M that we are hypothesizing as the generator of the data. Here we are interested in a different task - could we measure the difference between two quantities, $x - y$, even if we do not know what x and y are? Or could we measure the error in x without knowing x itself?

The comparison between this work and identifiability of causal quantities is more illuminating when we consider the mathematics used for either. In both, we are creating a linear system. One goal of identifiability is to identify what conditions will make the linear system a determined one - the existence of point estimates. Here we also establish a linear system but it will always be under-determined. Furthermore, it is not able to tell us what the counterfactual quantities are. Instead it solves for their differences. Our claim is that these systems, while under-determined, are not trivial and contain information about the experimental setting of a theory. Here we use it to allow for a purely experimental verification of the counterfactual equality assumptions in Cinelli and Pearl [1]. As we will see, only certain sets of equality can be relaxed and still be consistent with the data of the Aceh, West Java and Sarlahi Vitamin A trials.

The difference between GTI and identifiability is one of changing what we assume to be correct and valid. We could view data as an imperfect sample from a well-defined SCM we assume to be correct. Or we could view data as sacrosant and use it to tell us about what range of theories could explain it.

Sensitivity analysis is also another way to take into account the uncertainty we may have about all the properties of a model. Like Cinelli and Pearl [1], we compute some estimate for a causal quantity but aware that it could depend sensitively on the assumptions simplifications we made to arrive at the result. In addition, we posit some parametrization of deviations from the assumptions and then derive functions that are able to give us a measure of the sensitivity of our conclusions to the assumptions we made. An example is given in Cinelli and Pearl (2018). Here we also derive what could be construed as a sensitivity analysis of the counterfactual assumptions used by Cinelli and Pearl [1].

Algebraic statistics has been used before in CI by García-Puente and Sullivant [6]. Intellectually speaking, algebraic GTI should be considered as part of Algebraic Statistics. Historically, all textbooks so far published are speaking about Algebraic Parametric Statistics, whereas Algebraic GTI is Algebraic Non-parametric Statistics. No generalization can be made from the calculations performed with Algebraic GTI since it is non-parametric.

2 The Counterfactual Equality Assumptions

Cinelli and Pearl [1] are concerned with the generalization of Random Controlled Trials (RCTs) across domains. To do this, they formulate this theorem (theorem 2 in their paper) that assumes the equality of two counterfactual quantities across domains - the probability of sufficiency for the binary outcomes of the experiments they consider (PS_{01} and PS_{10}). The probabilities of sufficiency are counterfactual quantities. As vividly portrayed by Cinelli and Pearl [1] with their example of the health effects of playing Russian Roulette, sometimes we have scientific reasons for believing that an effect is so strong as to be valid across different domains. In the context of the extreme example of playing Russian Roulette, the first probability of sufficiency ,PS_{01}, is the share of people that would have died if they were forced to play among those that would not have died if not forced to do so. The second probability of sufficiency, PS_{10}, is the probability that playing Russian Roulette is sufficient to save a person who would have died if not forced to play.

As argued by Cinelli and Pearl [1], their causal diagram leads them to assert that these probabilities of causation are invariant across domains. They use this in the following theorem (numbered 2 in their paper),

Theorem 1. *Consider two source domains Π^a and Π^b. Let the probabilities of sufficiency be the same across the two populations, that is, $PS^a_{01} = PS^b_{01} = PS_{01}$ and $PS^a_{10} = PS^b_{10} = PS_{10}$. Then,*

$$PS_{10} = 1 - \frac{P^a_{11}P^b_{00} - P^b_{11}P^a_{00}}{P^a_{01}P^b_{00} - P^b_{01}P^a_{00}} \tag{1}$$

$$PS_{01} = 1 - \frac{P^b_{11}P^a_{01} - P^a_{11}P^b_{01}}{P^a_{01}P^b_{00} - P^b_{01}P^a_{00}}. \tag{2}$$

Where . . .

This allows them to generalize RCTs in two domains to a third one. Strict equality is rare in Science. And asserted, it becomes an object of scientific study requiring experimental verification. As an example, consider the equality between gravitational and inertial mass that has been verified experimentally by physicists down to 10^-13, Schlamminger et al. [7]. We now show one way to do a similar verification using the data in the Cinelli and Pearl [1] paper.

3 Under-determined linear algebra systems for verifying counterfactual assumptions across domains

To verify the equality assumptions for the probabilities of sufficiency across the domains, we will be constructing an under-determined system that eliminates all PS quantities and instead looks at their differences. The following combinatorial discussion shows why this approach may be interesting given many domains.

3.1 Combinatorial aspects of equalities

Consider a generic assumption of equality between three quantities,

$$a = b = c. \tag{3}$$

To be verified experimentally, we require verification of three equations

$$a = b \tag{4}$$

$$a = c \tag{5}$$

$$b = c \tag{6}$$

For the case here, assuming equality of the probabilities of sufficiency ,either PS_{01} or PS_{10}, across n domains would require

$$\frac{n(n-1)}{2} \tag{7}$$

equality assumptions. Since we have two such probabilities of sufficiency, we have $n(n-1)$ equalities. We find that the under-determined system constructed here is able to relax

$$\frac{n(n-1)}{2} - 1 \tag{8}$$

of these assumptions. At $n = 3$ this gain is paltry. Six equalities are assumed but we can only relax 2 to obtain a single measurable difference as shown in Table 1. But this changes as the number of domains increases, then the observational data allows for more detailed differences between the domains. We will discuss how these differences increase once we present the formulas.

Most surprising, however, is that the relaxation of assumptions cannot mix the types of probabilities of sufficiency - all relaxed equalities must be of the same type, either PS_{01} or PS_{10}. We do not know if this is a trivial result or not. Perhaps a reader more familiar with these causal quantities can provide clarification on this un-mixing property.

3.2 The general approach

Given n domains we construct a linear system that contains n equations of the following form for each domain d,

$$P_{11}^d = (1 - PS_{10}^d)P_{01}^d + PS_{01}^d P_{00}^d \tag{9}$$

In addition we have two equations for each pair of domains, d_i and d_j, that introduce the Δ variables that will allow us to relax the assumptions of equality.

$$\Delta_{01,d_1,d_2} = PS_{01}^{d_1} - PS_{01}^{d_2} \tag{10}$$

$$\Delta_{10,d_1,d_2} = PS_{10}^{d_1} - PS_{10}^{d_2} \tag{11}$$

A fast and loose argument shows that this system is under-determined. We have, by construction,

$$n + 2\frac{n(n-1)}{2} = n^2 \tag{12}$$

equations. But we have to solve for

$$2n + 2\frac{n(n-1)}{2} = n^2 + n \tag{13}$$

unknown variables,the Δs and the PSs. This accounting is loose because the Δ unknowns are not independent of each other. Given three domains, knowing the difference between two pairs of them, tells us the difference between the last pair.

The detailed calculation shows that, in fact, this system becomes determined when we assume that

$$\frac{n(n-1)}{2} + 1 \tag{14}$$

of the Δs are zero.

3.3 The $n = 3$ formulas

Let us now consider the $n = 3$ case. Straight forward linear algebra calculations show that if we assume that all of the PS_{10}^d are equal and just one pair (a,b) has equal PS_{01}, then the difference in PS_{01} with a third domain c and these two equal domains is given by,

$$\Delta_{01,b,c} = \frac{-P_{11}^a(P_{00}^b P_{01}^c - P_{00}^c P_{01}^b) + P_{11}^b(P_{00}^a P_{01}^c - P_{00}^c P_{01}^a) - P_{11}^c(P_{00}^a P_{01}^b - P_{00}^b P_{01}^a)}{P_{00}^c\left(P_{00}^a P_{01}^b - P_{00}^b P_{01}^a\right)} \tag{15}$$

and, of course, $\Delta_{01,b,c} = \Delta_{01,a,c}$.[1] This was the equation used to calculate the entries in Table 1.

A second verification of the assumption of equalities is also possible if we now assume that all of the PS_{01}^d are equal and just one pair (a, b) has equal PS_{10},

$$\Delta_{10,b,c} = \frac{P_{11}^a(P_{00}^b P_{01}^c - P_{00}^c P_{01}^b) - P_{11}^b(P_{00}^a P_{01}^c - P_{00}^c P_{01}^a) + P_{11}^c(P_{00}^a P_{01}^b - P_{00}^b P_{01}^a)}{P_{01}^c\left(P_{00}^a P_{01}^b - P_{00}^b P_{01}^a\right)} \tag{16}$$

Providing us with Table 2

[1] Care must be taken with the sign of the delta variables since it depends on the ordering of the domains used to define a difference. So using the lexicographic ordering of domains (a,b,c), we would have $\Delta_{\cdot,b,c} = -\Delta_{\cdot,c,b}$

pair	Δ_{10}
$\Delta_{10,\text{Aceh,West Java}} = 0$	-0.0021
$\Delta_{10,\text{Aceh,Sarlahi}} = 0$	-0.0027
$\Delta_{10,\text{Sarlahi,West Java}} = 0$	-0.0082

Table 2: Deviations of PS_{10} counterfactual assumptions from equality assuming all PS_{10} counterfactuals are equal

3.4 The $n > 3$ formulas

Similar equations are obtained for the case $n > 3$ whenever we are computing a difference between a third domain and either of the two domains we are assuming are equal in their probabilities of sufficiency. But now a second comparison type becomes possible - between two domains other than the two assumed to be equal. This gives us two new equation types. For simplicity, we assume that the domains are (a,b,c,d) so we can keep track of the signs for the delta variables. Assume that all the PS_{10} are equal but only one PS_{01} pair, the (a,b) pair, is equal. We then obtain for any other pair (c,d) the following experimental discrepancy,

$$\begin{aligned}\Delta_{01,c,d} = (&-P^d_{00}P^c_{11}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right) + P^b_{00}P^a_{11}\left(P^c_{00}P^d_{01} - P^d_{00}P^c_{01}\right) - \\ &P^a_{00}P^b_{11}\left(P^c_{00}P^d_{01} - P^d_{00}P^c_{01}\right) + P^c_{00}P^d_{11}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right))/ \\ &\left(P^c_{00}P^d_{00}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right)\right). \end{aligned} \tag{17}$$

Likewise, assuming that all PS_{01} are equal, and just the (a,b) pair PS_{10}s are equal gives us the following equation for any other (c,d) pair,

$$\begin{aligned}\Delta_{10,c,d} = (&-P^d_{01}P^c_{11}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right) + P^b_{01}P^a_{11}\left(P^c_{00}P^d_{01} - P^d_{00}P^c_{01}\right) - \\ &P^a_{01}P^b_{11}\left(P^c_{00}P^d_{01} - P^d_{00}P^c_{01}\right) + P^c_{01}P^d_{11}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right))/ \\ &\left(P^c_{01}P^d_{01}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right)\right) \end{aligned} \tag{18}$$

4 Mixed relaxations are zero

So far we have presented results where all of either the PS_{01} or the PS_{10} are assumed to be equal across all domains. This allowed us to relax all but one of the other pair of equalities for the other counterfactual type. Consider now what happens when we mix these equalities. In particular, consider the case of $n = 3$ domains. We saw that we could allow two of the equalities to relax and they both had to have the same type. But suppose we now assume that only one of each type is allowed to relax, say, $\Delta_{01,a,b} \neq 0$ and $\Delta_{10,b,c} \neq 0$. The only way this can solve the resulting system of linear equations is if the following two equations hold,

$$-P^a_{11}\left(P^c_{00}P^b_{01} - P^b_{00}P^c_{01}\right) - P^b_{11}\left(P^a_{00}P^c_{01} - P^c_{00}P^a_{01}\right) + P^c_{11}\left(P^a_{00}P^b_{01} - P^b_{00}P^a_{01}\right) = 0 \tag{19}$$

$$\Delta_{01,a,b} = 0 \tag{20}$$

$$\Delta_{10,b,c} = 0 \tag{21}$$

Using the Vitamin A trials data from Cinelli and Pearl [1], the equality 19 is, in fact, nearly true. The left hand side evaluates to,

$$\frac{6167349400}{200083835457} + \frac{54049356}{13295735299} - \frac{6721394800}{192613371687} \tag{22}$$

That such different integer ratios should resolve to nearly zero ,-0.000007, seems an extremely fortuitous result. Why is the assumption that two domains are equal in PS_{01} and the same or another pairs is equal in PS_{10} allow us to confirm experimentally that all probabilities of sufficiency are equal across all domains? This can be re-stated as the following theorem,

Theorem 2. *For any domains a,b and c if there is a pair where the PS_{01} and PS_{10} are equal, then, if Equation 19 holds, all probabilities of sufficiency are equal across all three domains.*

5 Conclusions

We have presented a series of calculations that allow us to use the experimental data for three domains to measure the tightness of counterfactual quantities, the probabilities of sufficiency, across the domains. In addition, we find that experimental data can sometimes allow us to relax assumptions and still discover that the equality must be strict.

This "toy application" of algebraic GTI is intended as a first step to solve the more interesting problem of how we can measure the independence of U errors in a non-parametric fashion.

References

[1] Carlos Cinelli and Judea Pearl. Generalizing experimental results by leveraging knowledge of mechanisms. *European Journal of Epidemiology*, page to appear.

[2] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the Natural Academy of the Sciences*, 113(27):7345–7352, 2016.

[3] Jonathan G. Richens, Ciarán M. Lee, and Saurabh Johri. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(3923), 2020.

[4] Hasok Chang. *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, New York, 2004.

[5] Andrés Corrada-Emmanuel, Edward Pantridge, Edward Zahrebelski, Aditya Chaganti, and Simeon Simeonov. Algebraic ground truth inference: Non-parametric estimation of sample errors by ai algorithms, 2020. URL `https://arxiv.org/abs/2006.08312`.

[6] Spielvogel S. García-Puente, L. D. and S. Sullivant. Identifying causal effects with computer algebra. In *Proceed-ings of the Twenty-Sixth Conference on Uncertainty inArtificial Intelligence (UAI)*. 2010.

[7] S. Schlamminger, K.-Y. Choi, T. A. Wagner, J. H. Gundlach, and E. G. Adelberger. Test of the equivalence principle using a rotating torsion balance. *Physical Review Letters*, 100(4), Jan 2008.