Causal Graphs and Sources of Bias

David Meyer

dmm@{1-4-5.net,uoregon.edu,...}

Last update: December 15, 2018

1 Introduction

Inferring the causal structure of a set of random variables from a finite sample of the joint distribution is an important problem in science. One way to think about causal structure is that when we observe the world, we are really observing some joint distribution $P(X_1, \ldots, X_n)$. However, we don't really know which of the X_i's are "causes" and which are "effects" other than that some set of the X_i's is the effected variable Y; the X_i's are sometimes called "treatment variables" and Y is called the "response variable".

Unfortunately these cause/effect relationships aren't represented in non-interventional data. To understand these relationships we need what is called a Markov (or causal) graph. A graph G is called Markov if it is a DAG and for a given node in the graph, that node is conditionally independent of its non-descendants given its parents. That is, $x_j \perp\!\!\!\perp nd_j \mid pa_j$, where nd_j are the non-descendants of x_j in G and pa_j are the parents of x_j in G. Said another way, the joint distribution $p(x_1, \ldots, x_n)$ factorizes:

$$p(x_1, \ldots, x_n) = \prod_{j=1}^{n} p(x_j \mid pa_j)$$

The semantics are that the parents pa_j of x_j in G are x_j's direct causes.

A perhaps more abstract way to think about the Markov condition is that given all the direct causes of an observable O, its non-effects provide no additional information on O [1]. The underlying idea here is that "nature chooses" the conditional distributions $p(x_j \mid pa_j)$ independently from one another, since the generation of additional independencies (that is, independencies are not imposed by the structure of the DAG) would require tuning between these conditional distributions [1].

If we do have a such a graph G, can we say anything about the associations (statistical relationships) implied by G? We might think about these associations, that is, the statistical relationships that we observe, more as a bundle of different kinds of associations, some of which may be causal and others which are spurious (non-causal). The goal of *causal identification* is to eliminate all the non-causal associations (which turn out to be paths in G), leaving only the causal relationships, if any. Figure 1 is a cartoon of this situation.

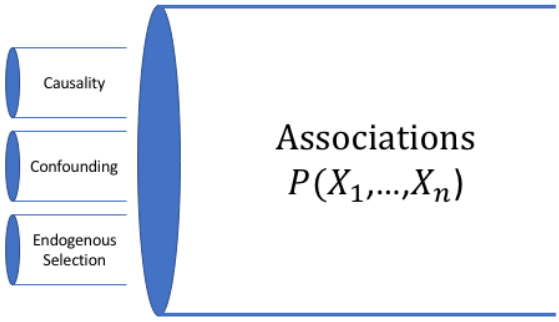

Figure 1: Associations, joint distributions, and bais

2 Causality

Causality in a Markov graph is represented by a chain structure in which all the arrows are pointing in the same direction (the direction of "time"); this is shown in Figure 2. A bit of notation: Two real-valued random variables A and B are said to be *conditionally independent* given C, written $A \perp\!\!\!\perp B \mid C$, if $\forall A, B\ P(A, B \mid C) = P(A \mid C)\, P(B \mid C)$ and for $\forall C\ P(C) > 0$. Note that conditional independence neither implies nor is implied by independence. That is, there are A, B and C such that we have only independence or only conditional independence.

In our example, we have $A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B \mid C$, that is, A is not independent of B (because A causes B), but conditioning[1] on C renders A independent of B. This is called *Overselection Bias* because you conditioned on a node in the chain, resulting in independence ($A \perp\!\!\!\perp B \mid C$) when in reality A and B are causally related ($A \not\perp\!\!\!\perp B$). The solution to Overslection Bias is not to condition on a node in the chain (here C).

[1]Conditioning can be as simple as marginalizing out C.

3 Confounding

A *confounder* is a (possibly hidden) common cause in a Markov graph. This is shown in Figure 3. The basic idea here is that there is non-causal path in the graph that is transmitting an association. Here A seems to be causally related to B ($A \not\perp\!\!\!\perp B$) due to this association. However, if we remove the non-causal path by conditioning on the confounder C, we see that A doesn't really cause B ($A \perp\!\!\!\perp B \mid C$). This bias is called *Confounding Bias* or sometimes just Confounding. A general solution to Confounding is to condition on the common cause (C). Essentially Confounding bias arises from the failure to condition on a common cause.

One of the most famous examples of confounding involves the eminent statistician R.A. Fisher, who claimed (among other things) that the presence of a confounder meant that one couldn't conclude that there was a causal link between smoking and lung cancer [2]. This is shown in Figure 4.

4 Endogenous Selection

The structure of the graph representing Endogenous Selection is shown in Figure 5. You can see how *Endogenous Selection Bias* comes about in the following classic example [3]: Consider the following causal model for the relationships between productivity, A, originality, B, and academic tenure, C. C here is called a "collider". Now, suppose that productivity and originality are unassociated in the general population (i.e., productivity does not cause originality and originality does not cause productivity (i.e. $A \perp\!\!\!\perp B$), and productivity and originality do not share any common cause). Suppose further that originality and productivity are separately sufficient for promotion to tenure. In this case, tenure is a collider variable.

Now if you condition on tenure (the collider, C) then you are assessing the relationship between originality and productivity only among tenured faculty (showing that Endogenous Selection Bias subsumes Sample Selection Bias [4]). In this case knowing that an unoriginal scholar has tenure implies that he must have been productive. Conversely, knowing that an unproductive scholar has tenure implies that he must have been original. In either case conditioning on the collider tenure (C) creates an association between productivity (A) and originality (B) among tenured faculty, even though one does not cause the other. This form of bias is called *Endogenous Selection Bias* and is closely related to Berkson's paradox [5] and "explaining away" [6].

Perhaps surprisingly, in Endogenous Selection Bias A and B are independent (A and B don't cause one another, $A \perp\!\!\!\perp B$), but conditioning on (knowing) C induces a relationship between A and B

Figure 2: Causality (chain): $A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B \mid C$

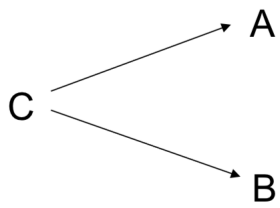

Figure 3: Confounding (common cause): $A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B \mid C$

that doesn't exist in the population. That is, $A \not\perp\!\!\!\perp B \mid C$. The solution here is not to condition on a collider (or any node downstream of the collider). Endogenous selection bias results from the mistaken conditioning on a common effect.

5 d-separation

Finally, can you read any of these relationships directly off the graph? In certain situations the answer is yes. The rule we use for this, called *d-separation*, is due to Pearl [7]. Pearl describes d-separation as a "gift from the gods" since it allows us to deduce properties of the statistical distribution (associations) from the causal graph. Of course, in order to do this, you need the causal graph. So where does the causal graph come from? That is a whole different issue, but suffice it to say that nature doesn't reveal causal structure in non-interventional data, so we have to get it somewhere else. Discovering the causal graph (or other representation) is called *causal discovery*. In causal discovery (also called structure learning) we're trying to reconstruct the structural causal model or its graphical representation from its joint distribution $p(X_1, \ldots, X_n)$.

In general, we would like to ask the question: Under what assumptions on the data generating

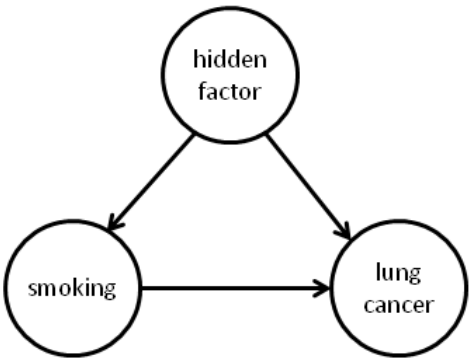

Figure 4: R. A. Fisher, smoking, lung cancer and confounding

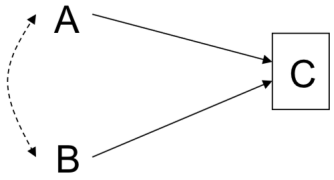

Figure 5: Endogenous Selection (common outcome): $A \perp\!\!\!\perp B$ and $A \not\perp\!\!\!\perp B \mid C$

process can one infer the causal graph from the joint distribution? There is pretty vast literature here. One place to start might be [8].

References

[1] J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf, "Causal discovery with continuous additive noise models," *J. Mach. Learn. Res.*, vol. 15, pp. 2009–2053, Jan. 2014.

[2] R. A. FISHER, "Cancer and smoking," *Nature*, vol. 182, pp. 596–, Aug 1958.

[3] F. Elwert and C. Winship, "Endogenous selection bias: The problem of conditioning on a collider variable," *Annual Review of Sociology*, vol. 40, no. 1, pp. 31–53, 2014.

[4] C. Cortes, M. Mohri, M. Riley, and A. Rostamizadeh, "Sample Selection Bias Correction Theory," *ArXiv e-prints*, May 2008.

[5] D. Westreich, "Berkson's bias, selection bias, and missing data," *Epidemiology*, vol. 23, pp. 159–164, Jan 2012. 22081062[pmid].

[6] M. P. Wellman and M. Henrion, "Explaining 'explaining away'," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, pp. 287–292, Mar. 1993.

[7] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.

[8] J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf, "Identifiability of causal graphs using functional models," in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI'11, (Arlington, Virginia, United States), pp. 589–598, AUAI Press, 2011.