A review of nested sampling

Abstract

Causal inference is a often-negleted topic from standard statistical textbooks. Historically, there have been two frameworks for thinking about causality: the potential outcomes framework, developed by Neyman and Rubin from 1923 to 1950s, and the structural framework, introduced by Judea Peal. In this short document, we introduce the structural causal model, causal effects, counterfactual analysis and the do operator. Also, we show how Neyman's potential outcome framework may be interpreted as a special case of the Pearl's structural equations.

1. Causal vs. Associational

Causal questions are questions that cannot be answered from the joint distribution of the observed variables alone. Properties that can be derived exclusively from the calculus of probabilities are associational in nature and do not require further theoretical development. Randomization, influence, effect and "holding constant" are examples of causal concepts; "controlling for", conditionalization and likelihood are examples of associational concepts.

2. Structural causal models

To deal with causal queries, we must extend the syntax of probability calculus. Figure 1 shows an example of a structural model; there, U_X and U_Y are exogeneous variables which cannot be directly measured, and X and Y are the endogeneous variables. When U_X and U_Y are not uncorrelated, we may estimate the direct effect β by $\beta = \mathrm{Cov}(X, Y)$ (we are assuming that variables are centered with unit variance).
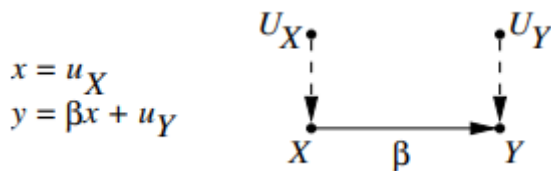

$$x = u_X$$
$$y = \beta x + u_Y$$

Figure 1: Example of structural model.

A critical definition for embedding a SCM with a notion of directionality is that of d-separation. We say that a set of nodes S d-separates variables X and Y if either for every path between X and Y (i) there is an arrow-emitting node within this path that is a member of S or (ii) there is a collider within this path that is not a member of S and has no descendants of S. In particular, if there is a collider between X and Y, then the empty set d-separates the corresponding path.

3. Interventions

Figure 2 shows an example of an intervention, which is the structural equivalent of holding a variable constant. For a structural equation model

$$z = f_Z(u_Z),$$
$$x = f_X(z, u_X), \qquad (1)$$
$$y = f_Y(x, u_Y),$$

it corresponds to letting $x = x_o$ and removing all incoming arrows to X, thereby defining a novel probabilistic model M_x instead of the non-interventional M characterized by the equations above; the operation of holding x constant is denoted by $\mathrm{do}(x = x_o)$. Under these circumstances, the problem of identifiability boils down to checking whether a quantity Q (e.g., $\Pr[y \mid \mathrm{do}(x)])$ is uniquely defined by the probability distribution induced by the intervened model, i.e., $\Pr[M_1] = \Pr[M_2] \Rightarrow Q(M_1) = Q(M_2)$.
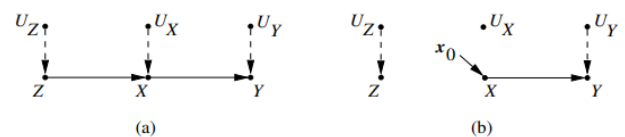

(a)   (b)

Figure 2: Example of an intervention on a variable X.

For example, under the model above, $\mathbb{E}[Y \mid \mathrm{do}(X = x_o)]$ is identifiable. Similarly, any quantity reliying exclusively on a endogeneous variables is identifiable by the Causal Markov Condition, i.e.,

$$\Pr(v_1, ..., v_n) = \prod_i P(v_i \mid \mathrm{pa}_i) \qquad (2)$$

in which v_i are endogeneous variables and pa_i is the set of parents of v_i.

Importantly, when dealing with unmeasured confounding, the back-door criterion may be employed to identify an admissible (sufficient) set of measured variables that ensure the identifiability of an effect. In a nutshell, the **back-door** criterion is defined for a set S and variables X and Y by

1. no node in S is a descendant of X;
2. S blocks (i.e., d-separates) every back-door paths from X to Y, namely, a path ending with an arrow pointing to X.

When a sufficient set S is a available, we may write $\Pr[Y = y \mid \mathrm{do}(X = x), S = s] = \Pr[Y = y \mid X = x, S = s]$. Generally, such a set S may not exist and we would have to rely on multi-stage approaches that recursively carry out inference on increasing subgraphs. In Figure 3, for instance, when only X, W_3 and Y are measurable, there is no admissible set but we can first estimate $\Pr[w_3 \mid \mathrm{do}(x)]$ (identifiable) and then $\Pr[y \mid \mathrm{do}(w_3)]$ (identifiable) and then $\Pr[y \mid \mathrm{do}(x)]$ with the rules of probability.
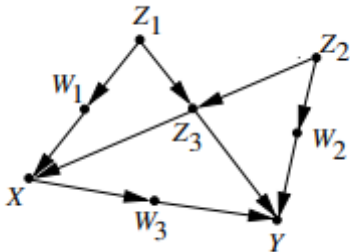

Figure 3: Example of an intervention on a variable X.

4. Counterfactual Analysis and Causal Effects

A (unit-level) counterfactual refers to the value of a target variable Y had a treatment variable X assumed a value x, i.e., $\mathrm{do}(X = x)$. We denote this quantity by $Y_{x(u)} = Y_{M_x}(u)$, in which the latter is the random variable corresponding to Y in the intervened model for which $X = x$. As we will see next, this approach naturally leads to Neyman's potential outcome framework.

Under these conditions, a causal effect is measured as the difference in expectation between a target variable given the operated treatment and given a counterfactual treatment. Namely,

$$\Pr[y \mid \mathrm{do}(x)] - \Pr[y \mid \mathrm{do}(x')]. \qquad (3)$$

Generally, these quantities are not idenfiable and the best we can do is to find a lower- and upper-bound for them. There is a rich literature concerning this problem, which we do not cover here.

5. Potential outcomes

Neyman-Rubin's framework of potential outcomes comes from a mathematical tradition of defining a primitive of interest and deriving whatever can be derived from this definition and related results alone.

More specifically, Neyman-Rubin's axiomatic approach starts by defining the hypothetical quantity, $Y_{x(u)}$, and considering the observed distribution as the marginal distribution of an augmented probability that includes both the observed and counterfactual variables. In this scenario, the counterfactual analysis becomes a problem of missing data, which may be dealt with any tool from the whole utility belt of frequentist statistics. These novel variables, however, are not whimsy; they are restricted by consistency constraints such as $X = x \Rightarrow Y_x = Y$. Pearl shows that such definitions may be derived from the structural model paradigm, and advocates for the use of the latter in opposition to the former. Indeed,

in spite of its mathematical soundness, it is scientifically more difficult to assure that the selected model satisfies the cognitively demanding axioms of the potential outcome framework [1].

Bibliography

[1] J. Pearl, "Causal inference in statistics: An overview," 2009.