Review of Pearl's Causal inference in Statistics

By Carlos Rubio

Pearl represents the *data generating process* (DGP) in one of two forms: using a **structural causal model** (SCP), or through **direted acyclic graphs** (DAG).

The **structural causal model** is a model developed by Pearl himself, though many of its basic ideas were introduced earlier, like the Wright's *structural equation models* (SEM) and Rubin's potential outcome framework. The structural equation models developed by Wright complemented the use of equations to express the effect a variable had on another with *DAGs*, were the incidence of the edges represented the direction of causality ($X \to Y$). The idea was that by conditioning on a subset of the nodes in the graph S, it was possible to achieve conditional independence from a node $Z \to X$, such that only the treatment X has an effect on the post-treatment variable Y, thus removing the noise caused by Z on the effects of $X \to Y$. This is notated as $Z \perp\!\!\!\perp Y \mid S$.

The next idea was to strip the equations of their algebraic form, thus introducing the idea of structural invariance: changing the algebraic form of one of the functions that defines the variables does not change the form of the others. Structural invariance allows for intervention $do(x)$, where one of the functions of the model (namely, the treatment X) is changed by a constant x. Given the directionality of causality, this will help estimate the effect of X over Y.

The most important concept is the definition of admisible sets of unconditioned covariates in order to calculate the effect of a treatment on a variable: the back-door criterion. Colloquially, every node that is not a mediator (descendant of the treatment X that has an effect on Y) must be conditioned.