Carson Lindholm

Statistics DRP

13 December 2024

Causal Inference with Spatial Cofounding: Final Report

This quarter, we read through the paper *A causal inference framework for spatial confounding* to learn about general causal inference and the special case of the cofounders being meaningfully distributed over space. We first went through the main assumptions of causal inference that make our calculated shift estimand or average treatment effect rigorous. The main assumptions are consistency, positivity, and shift ignorability. Consistency means that the observed outcome is the counterfactual outcome under the observed treatment, so there are not multiple forms of treatment that lead to an outcome, nor does one's treatment influence others' outcomes. Positivity means there is a non-zero probability that any treatment group will receive a treatment (informally this means that all groups of people have a good mix of treatment levels). Shift ignorability means that our causal analysis considers all confounding variables and that nothing is left out of the analysis that might make the results unreliable. With these assumptions met, we can validly calculate the average difference in treatment outcomes $E[Y(x) - Y(x')]$. In spatial causal inference, we extend these basic assumptions. Most notably, we consider a set of unknown confounders U that we do not have access to in our data. In this case, we then consider the function $U = g(S)$, which must be a smooth function of our spatial information, usually latitude and longitude data. In spatial causal inference, we argue that the unknown confounders are completely representable by our function $g(S)$ so that even if we do not have access to the data, are spatial information is enough to meet shift ignitability. In the continuous setting, which most spatial causal inference falls under, after meeting the assumptions, we can then calculate the shift estimand $E[Y(x + \delta) - Y(x)]$ where δ is how much we shift our exposure variable X by. Our estimand represents the average increase or decrease in our outcome variable Y.

In the second half of our project, we investigated different algorithms to calculate our shift estimand in ways that best represent the real world. We mainly focussed on the difference between linear and nonlinear models. One main downfall of using linear models in spatial modeling is that shift estimand is constant across our space. A constant shift estimation is usually different from the real world, as shown in our working example of the effect of air pollution on cardiovascular mortality. Intuitively, we can understand that decreasing air pollution by a constant amount across the U.S. will not produce a constant decrease in the cardiovascular mortality rate. For this reason, we tend to prefer nonlinear algorithms in spatial modeling.

We finished our project by building some basic linear and nonlinear models. Some next steps for this project would be to focus more on building the nonlinear models and testing them to find a model that produces the most accurate shift estimand.