Krish Doshi
DRP Final Report

This quarter, I had the opportunity to explore Structural Causal Models (SCMs) with Weitao. Prior to this, I hadn't encountered SCMs, but I was eager to learn about a framework that blends statistics with fields like computer science. To prepare, I revisited key statistical concepts such as probability distributions, expectation, variance, covariance, and Bayes' Rule, which helped ground my understanding of causal modeling.

Early on, I learned that SCMs represent the world through variables and the causal relationships between them. Variables are split into two sets: exogenous (U), which are external and not explained by the model, and endogenous (V), which depend on other variables. A variable X is a direct cause of Y if it appears in the function determining Y's value. Interestingly, due to intransitivity, a direct cause might not always influence the final outcome.

As the course progressed, I studied the graphical structures that illustrate these relationships. I learned about product decomposition, where the joint probability of variables can be broken down into the product of each variable's probability conditioned on its parents. We explored patterns like chains, forks, and colliders. In chains, conditioning on the middle variable blocks dependence between the ends. Similarly, conditioning on the shared cause in a fork removes dependency. Colliders behave differently—without conditioning, the parent variables are independent, but conditioning on the collider introduces dependence. This is important in real-world scenarios like medical studies, where selection bias can lead to misleading conclusions.

Later, I explored intervention and the backdoor and front-door criteria for identifying causal effects. Intervening (using do(X)) sets a variable's value while cutting off its incoming influences, simulating a randomized experiment. In contrast, conditioning only filters observations based on a variable's value. The backdoor criterion helps block non-causal paths by conditioning on variables that d-separate the treatment and outcome, eliminating confounders. For example, controlling for age allows us to isolate the causal effect of a drug on recovery.

When confounders are unobservable, the front-door criterion provides an alternative. By identifying a mediator that fully transmits the causal effect, we can estimate the relationship even with unmeasured confounding. This is useful in cases like studying pollution's impact on health, where socioeconomic factors might be unobserved.

Throughout the quarter, I also explored real-world applications of SCMs, including their use in astronomy and potential in machine learning. I'm excited to continue building on this foundation and applying these concepts in future work.