Estimating Causal Effect

Mentee: Casey Logan

Mentor: Pawel Morzywolek

This quarter the topic that I learned about was causal effect. I learned about the difficulties encountered in estimating causal effect, the three important causal assumptions you need, and three different methods for estimating causal effect. Combining all of these things into one analysis, I also used these processes to estimate the causal effect in sample data. In the example, the data was taken from an observational study from 1997. The data was collected at Ohio Heart Health and was recorded by staff of the Lindner Center. The purpose of the study was to identify if, in patients who recently received a heart transplant, receiving an extra augmented therapy gene called ABCIX would have an impact on survival rates after six months.

To estimate causal effect in an ideal world, you would look at the effect that receiving both treatment and not treatment have in the same individual. These effects are known as counterfactual outcomes. The counterfactual outcome Y^1 is the outcome you would observe if an individual received treatment, and the counterfactual outcome Y^0 is the outcome you would observe if that same individual does not receive treatment. If you observe both of these outcomes on the same individual, then estimating causal effect is simple because any change in the outcome would be due to the change in the treatment variable. However, you are not often able to do this. For instance, in our example, you would not be able to observe the effect that both receiving treatment and not receiving treatment have on the same individual since it does not make sense for an individual to both receive extra therapy and not receive extra therapy at the same time.

The next best thing would be to use a randomized trial, where you randomly assign individuals to treatment groups. This is an ideal scenario to establish causal effect because, due

to randomization, the only difference between your treated and untreated groups is the treatment value. Any difference in the outcome will be due to the treatment, meaning the treatment caused an effect in the outcome. Again, randomization is not always an option. Sometimes performing a randomization trial is not possible and often it is unethical.

Since the first two ideal scenarios are not feasible, the scenario you would be looking at is an observational study. In an observational study, you would observe both the treatment value and the outcome value of an individual. A problem that you encounter in this scenario is an issue called confounding, which prevents you from being able to establish a causal relationship. Confounding is when you have a variable other than your treatment value affecting your outcome, so when you observe a change in the outcome you are not sure if this change is due to the treatment variable or the confounding variable.

Because of these difficulties, you need three causal assumptions to hold in order to establish causal effect. The first of these assumptions is consistency. The consistency assumption states that every individual who is receiving treatment is receiving the same exact treatment in the same exact way. And every individual who is not receiving treatment is not receiving treatment in the same exact way. The second assumption is exchangeability. The exchangeability assumption states that the two treatment groups are identical in every aspect except for their treatment value. This is the assumption that is important to prevent confounding. The third assumption is positivity. The positivity assumption states that every individual has a probability greater than zero of receiving treatment or not receiving treatment. If these three assumptions do not hold, you cannot interpret the effect causally. Even if you see an association between the treatment and the outcome, if these assumptions do not hold, that association cannot be identified as a causal association.

There are many different methods you can use to measure the causal effect. The one that we used most often this quarter, and the one used in our example, is the risk difference.

The risk difference measures the expected value of the difference between the counterfactual outcomes, $E[Y^1 - Y^0]$. When this difference is 0, there is no causal effect because the outcome is the same whether the individual is being treated or not. In our example data, we found a risk difference of -0.036. This means that there was a 3.6% reduction in the risk of the death when a patient received the extra augment therapy as treatment.

The three methods that we have for estimating causal effect are outcome regression, inverse probability weighting (IPW), and doubly robust estimating. Outcome regression is a method that creates a model to predict the outcome Y based on the treatment value and covariates of the individual. The covariates of the individual are the other variables relating to the individual that you have information on, that are not the treatment or outcome variables. Using the model, you can average your predicted outcomes and use a causal effect measure such as the risk difference to estimate the causal effect. The second method for estimating causal effect is inverse probability weighting. This is a method that predicts the treatment value from the covariates. This creates a pseudo-population where the association between the covariates and the treatment value is removed, which adjusts for any confounding in your study. To calculate the inverse probability weights, an individual receives a value known as the propensity score, which represents their probability of receiving treatment based on their covariates. The individual is then weighted based on the inverse of that probability to create the inverse probability weight.

The final method for estimating causal effect is the doubly robust estimator. This method combines the previous two, outcome regression and IPW, into one model. Now, you predict the outcome Y based on the treatment A and covariates L, and weight by IPW. In this method, only one of the models needs to be correct for the doubly robust estimator to be accurate. For example, if the outcome regression model is incorrect, the doubly robust estimator is still accurate as long as the IPW is correct. When using a doubly robust estimator in our example,

we found a risk difference of -0.061. This means there was a 6.1% reduction in the risk of death in for individuals who received the extra augmented therapy. The motivation behind using the doubly robust estimator is that you would expect the confidence interval of your causal effect measure to be smaller than the confidence interval of your causal effect measure from the other methods. This is because you are using more information about every individual, so as the name implies, the estimate is more robust.

In conclusion, to estimate causal effect you need three causal assumptions to hold: consistency, exchangeability, and positivity. When these assumptions hold, you can use different measures such as the risk difference to estimate the causal effect. To estimate this measure, there are three methods of estimation, outcome regression, inverse probability weighting, and a doubly robust estimator which combines outcome regression and inverse probability weighting into one model. Using a doubly robust estimator you would expect to find a risk difference with a smaller confidence interval, giving you higher confidence that your estimate is correct.