Introduction to Modern Causal Inference

Alejandro Schuler Mark van der Laan

Table of contents

Preface

Goals and Approach

Our goal is to get people with any amount of experience in formal mathematics past undergraduate probability to be able to take a vaguely-defined scientific question, translate it into a formal statistical problem, and solve it by constructing a tool that is optimal for the job at hand. This book is not particularly original! There are many texts and courses that share the same philosophy and have substantial overlap in terms of the pedagogic approach and topics. In particular, we owe a lot to the biostatistics curriculum at UC Berkeley and to more encyclopedic texts like (Van der Laan, Rose, et al. 2011). Nonetheless, we think that the particular selection of topics and tone in this book will make the content accessible enough to new audiences that it's worthwhile to recombine the material in this way. Think of this book as just another open window into the exciting world of modern causal inference.

Philosophy

This book is rooted in the philosophy of modern causal inference. What sets this philosophy apart are the three following tenets:

1. The first is that for all practical purposes, **the point of statistics is causal inference**. Ultimately, we humans are concerned with how to make decisions under uncertainty that lead to the best outcome. These are fundamentally causal questions that ask *what if* we did A instead of B? The machinery of statistical inference is agnostic to causality, but that doesn't change the fact that our motivation in using it isn't. Therefore there is little point in shying away from causal claims because noncausal claims are not often of any practical utility.
2. That leads to the second, often misunderstood point, which is that **there is no such thing as a method for causal inference**. This isn't a failing of the statistics literature so much as a failing of the naive popularization of new ideas (e.g. blog posts, low-quality papers). The process of statistical inference (for point estimation) only cares about estimating a parameter of a probability distribution and quantifying the sampling distribution of that estimate. There is nothing *causal* about that, and many are surprised to learn that the algorithms used for *causal* inference are actually identical to those used to make noncausal statements. What makes an analysis causal has little if anything to

do with the estimator or algorithm used. It has everything to do with whether or not the researcher can satisfy certain non-testable assumptions about the process that generated the observed data. It's important to keep these things (statistical inference and causal identification) separated in your mind even though they must always work together.

3. The last and perhaps greatest shortcoming of traditional statistics pedagogy is that it generally does not teach you to ask the question that makes sense scientifically and then translate it into a statistical formalism. This is partly because, in the past, our analytic tools were quite limited and no progress could be made unless one imposed unrealistic assumptions or changed the question to fit the existing methods. Today, however, **we have powerful tools that can then translate a given statistical problem into an optimal method for estimation** in a wide variety of settings. It's not always totally automatic, but at a minimum it provides a clear way to think about what's better and what's worse. It liberates you from the route if-this-kind-of-data-then-this-method thinking that only makes for bad statistics and bad science.

Taken together, these three points give this movement a clear, unified, and increasingly popular perspective on causal inference. We have no doubt whatsoever that this is the paradigm that will come to dominate common practice in the next decades and century.

1 Inference and Statistics

2 Causality and Identification

3 Efficiency Theory

4 Building Efficient Estimators

References

Van der Laan, Mark J, Sherri Rose, et al. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Vol. 4. Springer.