

# **Introduction to Modern Causal Inference**

Alejandro Schuler

Mark van der Laan

# Table of contents

<b>Goals and Approach</b>	<b>3</b>
Philosophy . . . . .	3
Pedagogy . . . . .	4
Rigor with Fewer Prerequisites . . . . .	6
Core Concepts . . . . .	7
<b>Topics</b>	<b>9</b>
<b>Acknowledgements</b>	<b>11</b>
<b>1 Inference and Statistics</b>	<b>12</b>
<b>2 Causality and Identification</b>	<b>14</b>
<b>3 Efficiency Theory</b>	<b>15</b>
<b>4 Building Efficient Estimators</b>	<b>16</b>
<b>References</b>	<b>17</b>

# Goals and Approach

## ! Important

This book is a **work in-progress!** Everything about it is subject to change.

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 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.

## Pedagogy

The modern approach to causal inference is already well-established and discussed at length in many high-quality resources and courses. So why have yet another book going over the same material?

Well, there are many different kinds of students with different backgrounds for whom different presentation styles work better! We've chosen an approach for this particular text that focuses on the following principles:

1. **Rigor with fewer prerequisites.** Many students are turned off or intimidated by the amount of mathematical formality that is required to rigorously understand modern causal inference. While you don't actually need a lot of math to use and understand the ideas, you do need *some* math to be able to work independently in the field. Since one of our goals is for you to be able to *construct your own* efficient estimator for a never-before-seen problem, we have to think precisely and not hand-wave away rigor.

For better or worse, the math you need is built on several layers of prerequisites- real analysis, probability theory, functional analysis, and asymptotic statistics (more on that later). In an ideal world, we would all have taken these courses and be prepared for what's next, but in practice we know many people have gaps in their knowledge.

In this book we try to address that by providing a lot of foundational background information about these topics as they come up. We do our best to build up intuition for the component pieces of things before diving into implications. We lean heavily on figures and natural English prose to explain these concepts alongside the formal mathematical notation. At a minimum, we try to be very clear about what the foundations actually *are* so that if you don't understand something you at least know where to look.

2. **Core concepts.** Modern causal inference is a huge field. There are so many exciting applications and theoretical developments that it can be hard for students to find their way at first.

We've tried to keep things as short and concise as possible in this book by focusing on core ideas and a limited number of examples. The point of this is to emphasize what is most important and to allow you to learn the fundamentals without getting distracted. That means there are many topics that aren't covered in this book (see section below), but the idea is that this book will give you the tools and mindset you need in order to continue your learning productively.

3. **A welcoming voice.** Modern causal inference is fast-moving and intense. It's easy to feel like you don't belong or aren't good enough to participate. The agnostic, impersonal tone that is best for writing a clear academic paper doesn't always create the warm, emotional connection that is necessary for us to feel secure and ready to learn.

To deal with that problem, the voice we use throughout this book is informal and decidedly nonacademic. We write in the first person and address you, the reader, directly. Figures are hand-drawn and cartoonish and we have links rather than formal citations. We provide personal commentary alongside agnostic technical material. All of this is deliberate and meant to help you feel comfortable and guided through this world. You are not alone in learning!

Other books may share some or all of these principles to some extent so we're not claiming that what we've done here is entirely unique. We're trying to create as many inlets as possible for students with different lived experiences. If the approach we take here works for you, that's great!

If you find this book isn't for you, that's also great! Here is a list of other resources that cover some of the same ground:

- [\*Semiparametric Doubly Robust Targeted Double Machine Learning: A Review\*](#) (Kennedy, 2022)
- [\*Semiparametric Theory and Missing Data\*](#) (Tsiatis, 2006)
- [\*Targeted Learning\*](#) (Van der Laan et al., 2011)
- [\*Unified Methods for Censored Longitudinal Data and Causality\*](#) (Laan & Robins, 2003)
- [\*Causal Inference: What If?\*](#) (Hernán & Robins, 2010)

## Rigor with Fewer Prerequisites

Despite our best efforts, you will need some background material to get the most out of this book. The *hard* prerequisites listed are non-negotiable (this book will make no sense to you at all if you don't know what a probability distribution is), but the *soft* prerequisites are completely optional. We don't cover these topics here because there are already so many good resources to learn them and there's no need to reinvent the wheel. We provide the curated references below so you know where to go to fill in the holes in your understanding as you go along.

**Hard Prerequisites** You'll need to know probability theory at the undergraduate level. Random variables, expectation, conditional expectation, probability densities should be familiar to you, and ideally you have at least a vague understanding of what convergence in distribution means.

You also need undergraduate multivariable calculus and linear algebra: integrals, derivatives, vectors, etc. should all be well-worn tools at your disposal.

If you're shaky on these topics, we recommend [Khan Academy](#). We also encourage you to check out the real analysis and measure-theoretic probability resources listed below, which will provide you with a much more useful, conceptual understanding.

**Soft Prerequisites** The following topics aren't necessary to grasp the arc of the arguments we'll make, but a fully rigorous understanding isn't possible without them. As we go along, we'll provide parenthetical commentary to elucidate concepts from these subjects that you may not be familiar with. We suggest attempting to read and understand this book all the way through and taking notes on the places where you feel like you're missing something. This will help you prioritize your time if you choose to fill in the gaps. None of this material is beyond you, we promise. It just takes a little time to work through it.

These books and videos give what we think is the most self-contained treatment of their respective topics that is in line with the didactic approach taken in this book. They tend to have a more conversational tone, provide more background and intuition to the reader, and have plenty of worked examples. As always, however, there are many alternatives that may work better for you depending on your needs. These subjects should be tackled in the order they are listed here:

Field	Topics to Focus On	Resources
Intro to Real Analysis	sequences, limits, series, continuity, convergence (pointwise, uniform), derivative	<a href="#">Understanding Analysis</a> (Ch 1-6) (Abbott, 2015)

Field	Topics to Focus On	Resources
Measure-Theoretic Probability	definition and construction of measure, Lebesgue integral and its properties, Radon-Nikodym theorem, change-of-variables formula, modes of convergence for random variables, independence, central limit theorem, conditional expectation	<a href="#">Measure Theory YouTube Lectures</a> - Landim <a href="#">Measure, Integral, and Probability</a> (Capiński & Kopp, 2004)
Functional Analysis in $\mathcal{L}_2$	projection, bounded dual space, Reisz representation	<a href="#">Introductory Functional Analysis with Applications</a> (Ch 1-3) (Kreyszig, 1991)

A few tips for self study: we recommend reading skimming each chapter first and getting an idea of what the main ideas are and how they interconnect to each other and to what you've already done. Take your time to understand what the definitions of the objects are that are used in each theorem (e.g. what exactly is a sequence / random variable / linear bounded functional?). Think of some examples and draw pictures.

Then go through the chapter again and read it slowly, line by line. Reading a line of math should take you as long as it takes to read a paragraph of English because that's often how much information is conveyed therein. Now draw some pictures or think of examples attempting to illustrate the content of the theorems. After you're done, do several of the exercises that the author provides. We've found that it also helps to make a *one-page summary* or *cheatsheet* of all the material in the chapter (definitions, important theorems, and their relationships). This will help you keep all the concepts organized as you go along.

We'll also refer to [Asymptotic Statistics](#) (Van der Vaart, 2000) in certain places throughout the guide where we need to point to technical material. When we do this, we'll cite it as *vdV 98*. Relative to the way we write things up here, *vdV 98* goes over background material much faster, so you may find it more challenging to read through. However, it does contain all the necessary information and details. If you've read the material in the table above and you're feeling up to it you can try your hand at ch. 1, 2, 6, 7, 8, and 25 of that book.

## Core Concepts

We've done our best to keep this book lean and targeted on foundational concepts. As a consequence, there are many important and interesting topics that we'll say nothing or little about. For the sake of completeness, we'll list some of them here so you know where to start looking when you're ready to build on your solid foundation.

To keep things simple and consistent, most of the examples in this book focus on inference of the average treatment effect in simple covariate-treatment-outcome observational or randomized studies. However, there are many data structures which are much more elaborate than this. For example, we do not fully discuss inference in longitudinal settings with time-varying covariates and treatments, inference when some outcomes are censored, stochastic interventions, or continuous treatments. Non-IID data (e.g. treatment spillover in a network) is another interesting topic that is not covered. Nonetheless, what is in this book will prepare you to study all of these topics and to understand them as special cases or extensions of the general approach presented here.

While the tools we describe in this book are extremely powerful, they are not *all*-powerful. Some estimands of legitimate interest are not *pathwise differentiable* and therefore cannot be analyzed with these tools. An important example of such an estimand is the conditional average treatment effect, which is the subject of extensive research. There is not yet a unified approach to these problems.

Another interesting and useful setting that we do not discuss is online (or reinforcement) learning. Many topics related to this literature (e.g., off-policy evaluation) are directly addressable with the methods in this book, but here we omit all discussion of scenarios where the treatment policy can be manipulated during data collection as more information comes in. The framework for causal inference we develop in this book does extend to such settings but only with some added complexity that is best understood after learning the basics.

Lastly, the reader should be aware that this book discusses the most popular and widely used framework for causal inference, which is that of potential outcomes in the frequentist population model. Readers may have heard of directed acyclic graphs (DAGs) as an alternative to potential outcomes, but these two frameworks are in fact equivalent and easily unified (Richardson & Robins, 2013). Alternatives are exceedingly rare. Potential outcomes are also used outside the population model in the randomization inference framework, which differs in that only the treatment assignment is assumed to be random (i.e., sampling from a larger population is a source of variability). Much of what you learn here is also applicable to randomization inference. Potential outcomes are also commonly used in Bayesian approaches to causal inference, which we don't discuss here.



# Topics

This book aims to be a relatively self-contained treatment of the modern approach to causal inference. As we said above, the goal is for you to be able to take a vaguely-defined scientific question, translate it into a formal statistical problem, and solve it using a tool that you optimally construct for the job at hand. Thankfully, there's a well-developed process for doing that called the *Statistical Roadmap* (Petersen & Laan, 2014) which we follow closely in the chapters of this book:

1. To begin with, we need to first understand the process of *statistical inference*. We do that by defining a set of assumptions that represent what we know is true about the real world along with some true quantity of interest. Only then do we propose some kind of method or algorithm that, given data, produces an estimate of this quantity. We should be able to show that, with enough data, our answer gets closer and closer to the truth and that we have some way to quantify the uncertainty due to random sampling. These are the topics we'll work through in Chapter 1, closing out with an example that illustrates the danger of blindly applying popular methods without thinking through the question and assumptions.
2. Once we've understood the goal and process of inference, we introduce causality. Causal inference questions are just statistical inference questions but in an imaginary world where we can always observe the result of alternative *what if*'s. Of course in the real world we can't see all the *what if*'s. We only get to see what happened, not what *could* have happened had we made a different choice, taken a different drug, or implemented a different policy. To link the two worlds together we have to make some assumptions to guarantee that the answer to a statistical inference problem in the real world actually represents the answer to our causal question in the world of *what if*'s. Chapter 2 lays out this process (*causal identification*) and gives examples. Chapter 2 also addresses why a well-defined intervention is important to formulating something as a causal question and discusses broadly applicable methods that quantify the impact of potentially erroneous identifying assumptions.
3. Once we have an identification result in hand, we are completely done with causality and can focus on solving the real-world inference problem. But for better or for worse there are thousands of inference methods, each of which work under different assumptions and target different parameters. Choosing among these is daunting and often puts the cart before the horse because we're forced to define our question such that an existing method can solve it. Chapter 3 lays the foundation to address this problem by showing that for a wide range of assumptions and questions, there is generally a way to quantify what the

best possible estimator is in terms of the *efficient influence function*. The efficient influence function completely describes the behavior of the optimal estimator in large samples. This optimal estimator always converges to the right answer with more data and has the lowest legitimate uncertainty possible under the given assumptions. We work through examples of deriving efficient influence functions for the average treatment effect in standard observational and randomized experiments.

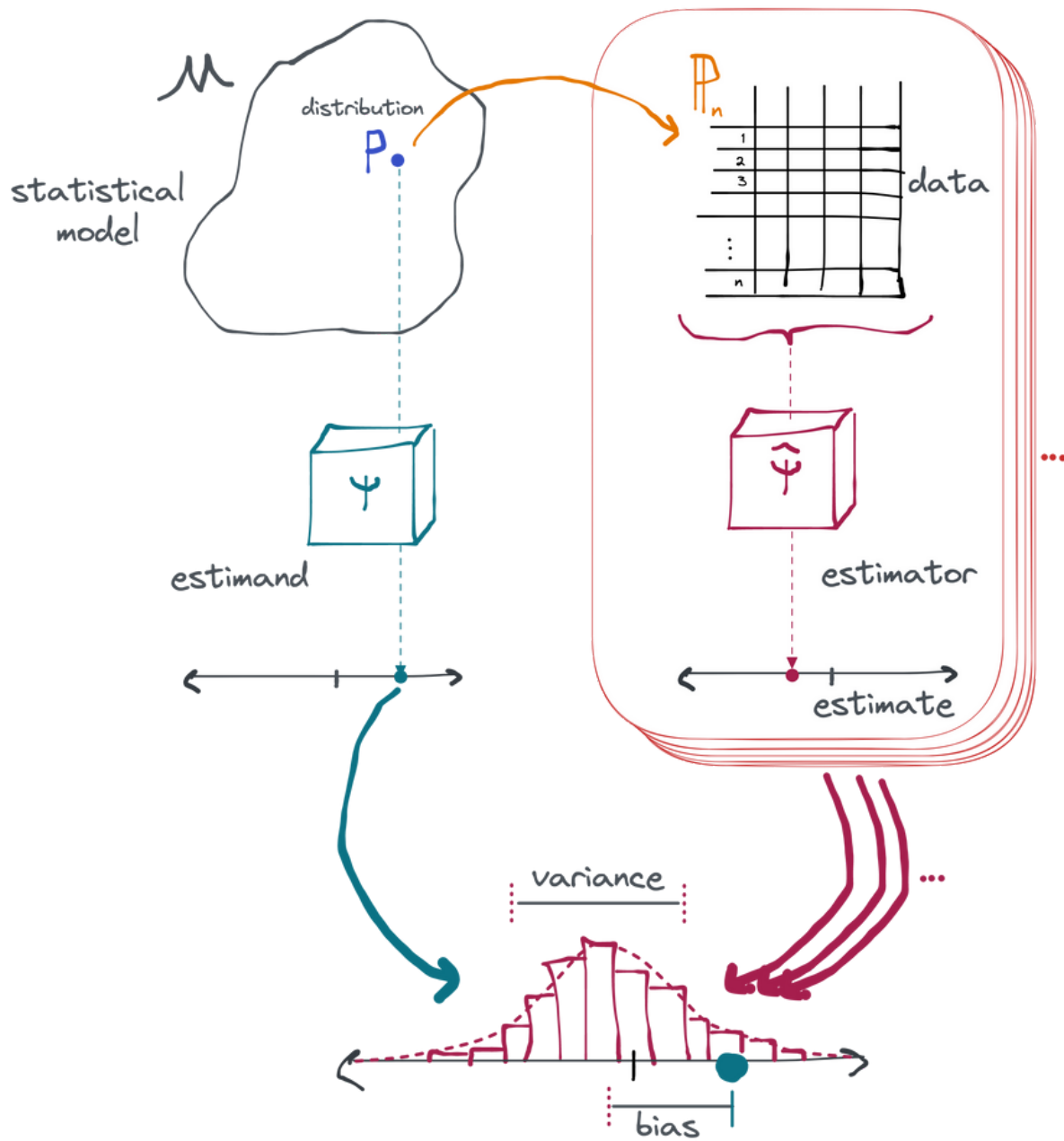
4. Finally, Chapter 4 does the hard work of putting the theory from Chapter 3 into practice. Here we show how we can actually build an *efficient estimator*- one which has the efficient influence function. In other words, given a statistical problem, we can construct an optimal method to solve it. We outline the three main strategies for doing this, which are called bias-correction, estimating equations, and targeted maximum likelihood. We show the different ways in which these strategies ensure that the constructed estimator has the right behavior. All of them ultimately attain the same goal and are theoretically equivalent in large-enough samples, but the differences are relevant for how the constructed estimators perform in practice.

# Acknowledgements

The authors thank Lauren Liao and Kat Hoffman for productive comments on chapter drafts.



# 1 Inference and Statistics



## 2 Causality and Identification

## 3 Efficiency Theory

# 4 Building Efficient Estimators



# References

- Abbott, S. (2015). *Understanding analysis*. Springer.
- Capiński, M., & Kopp, P. E. (2004). *Measure, integral and probability* (Vol. 14). Springer.
- Hernán, M. A., & Robins, J. M. (2010). *Causal inference*. CRC Boca Raton, FL.
- Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: A review. *arXiv Preprint arXiv:2203.06469*.
- Kreyszig, E. (1991). *Introductory functional analysis with applications* (Vol. 17). John Wiley & Sons.
- Laan, M. J., & Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer.
- Petersen, M. L., & Laan, M. J. van der. (2014). Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology (Cambridge, Mass.)*, 25(3), 418.
- Richardson, T. S., & Robins, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper, 128(30)*, 2013.
- Tsiatis, A. A. (2006). *Semiparametric theory and missing data*.
- Van der Laan, M. J., Rose, S., et al. (2011). *Targeted learning: Causal inference for observational and experimental data* (Vol. 4). Springer.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.