

Preface

This book describes a series of statistical methods for testing causal hypotheses using observational data – but it is not a statistics book. It describes a series of algorithms, derived from research in Artificial Intelligence, that can discover causal relationships from observational data – but it is not a book about Artificial Intelligence. It describes the logical and philosophical relationships between causality and probability distributions – but it is certainly not a book about the philosophy of statistics. Rather it is a *user's guide*, written for biologists, whose purpose is to allow the practising biologist to make use of these important new developments when causal questions can't be answered with randomised experiments.

I have written the book assuming that you have no previous training in these methods. If you have taken an introductory statistics course – even if it was longer ago than you want to acknowledge – and have managed to hold on to some of the basic notions of sampling and hypothesis testing using statistics, then you should be able to understand the material in this book. I recommend that you read each chapter through in its entirety, even if you don't feel that you have mastered all of the notions. This will at least give you a general feeling for the goals and vocabulary of each chapter. You can then go back and pay closer attention to the details.

The book is addressed to biologists, mostly because I am a practising biologist myself, but I hope that it will also be of interest to statisticians, scientists in other fields and even philosophers of science. I have not written the book as a textbook simply because the discipline to which the material in this book naturally belongs does not yet exist. Whatever the name eventually given to this new discipline, I firmly believe that it will exist, and be generally recognised as a distinct discipline, in the future. The questions that this new discipline addresses, and the elegance of its results, are too important. None the less, the chapters follow a logical progression that would be well suited to an upper level undergraduate, or graduate, course. I have used the manuscript of this book for such a purpose and every one of my students is still alive.

It is a pleasure and an honour to acknowledge the many people who have contributed to this project. First, Jim and Marg Shipley started everything. Robert van Hulst supplied much of the initial impulse through our conversations about science and causality while I was still an undergraduate. He has also read every one of the manuscript chapters and suggested many useful changes. Paul Keddy kept my interest burning during my Ph.D. studies and also commented on the first two chapters. As usual, his comments went to the heart of the matter.

The late Robert Peters had a large impact on my thoughts about causality and even convinced me, for a number of years, that ecologists are best to give up on the concept – not because he viewed the notion of causality as meaningless (he never believed this despite his empiricist reputation) but because it was simply too slippery a notion to demonstrate without randomised experiments. His constant prodding must have caused me to stop while wandering through the library one day when, almost subconsciously, I saw a book with the following provocative title: *Discovering causal structure. Artificial intelligence, philosophy of science, and statistical modeling* (Glymour *et al.* 1987). That book was my introduction to a more sophisticated understanding of causality. Rob Peters was much too young when he passed away and I am sorry that he never read the book that you are about to begin. I am not sure that he would have approved of everything in it but I know that he would have appreciated the effort.

Martin Lechowicz introduced me to the notion of path analysis at a time when this method had been mostly forgotten by biologists. He and I have collaborated for a number of years on this topic and he read the entire manuscript of this book, providing many insightful comments. Steve Côté and Jim Grace also read parts of this book. Jim, in particular, provided some important counterpoint to my thoughts on latent variable models. Marco Festa-Bianchet provided the unpublished data that is reported in Chapter 5. I must also acknowledge my graduate students, Margaret McKenna, Driss Meziane, Jarceline Almeida-Cortez, Luc St-Pierre and Muhaymina Sari, as well as the many members of the SEMNET Internet discussion group.

Finally, I want to thank Judea Pearl for kindly responding to my many emails about d-separation and basis sets and to Clark Glymour, Richard Scheines and Peter Spirtes of Carnegie-Mellon University for their generosity in extending an invitation to visit with them and for patiently answering my many questions about their discovery algorithms. Clark Glymour read and commented on some of the manuscript chapters.

I hope that you find this book to be useful, interesting and readable. I welcome your comments and feedback. Especially, if you don't agree with me.

Bill Shipley