Contents

Preface

If you believe in things that you don't understand, then you suffer (Wonder 1973).

When I was in graduate school in the early 1980s, linear model theory had just been perfected and I studied with some of the people who had perfected it. With its combination of simplicity and near-total explanatory power, linear model theory is like nothing else in statistics and by the early 1980s, this theory had been honed to the point where it could be taught almost whole to people taking their first serious regression course (e.g., Weisberg 1980).

At about the same time, however, the main thrust of statistical research turned in a different direction, emphasizing breadth over depth by producing methods for specifying and fitting models of greater generality and with weaker assumptions: generalized linear models, which dropped normality; additive models, which dropped linearity; generalized estimating equations (GEE), which dropped independence; hierarchical models, which added layers of structure; mixed models, which added random effects; Markov chain Monte Carlo, with the great flowering of Bayesian methods it enabled; the modeling syntax of the S and then R systems; structural equations models; spatial models and smoothers; dynamic linear models (state-space models); and penalized fits, among many others. Any statistician with a pulse has to love and be impressed by this explosion of sheer modeling power. I do, and I am. If you read something in this book that seems to contradict that statement, go back and read it again.

My admiration notwithstanding, I am mostly an applied statistician and when I use these new methods to analyze my collaborators' data, I routinely get results that are mysterious, inconvenient, or plainly wrong. For mixed linear models, these unhappy results include zero variance estimates, multiple maxima, counterintuitive outlier effects, odd fits (e.g., a wiggly smooth with one smoother but not with another apparently similar smoother), big changes in fit from apparently modest changes in the model or data, and unpredictable convergence of numerical routines, among other things. When my collaborators' datasets produce such puzzles, I urgently need something like the powerful theory of linear models so I can explain them and figure out what to do. As far as I can tell, however, little if anything is known about most of these puzzles. I see very few mentions of them in the listservs I peruse, in the statistical literature, or in talks. There's no reason to think I am a magnet for freak problems, so I suspect that many, perhaps most, working statisticians encounter the same puzzles. When I publish a paper or do a talk about some of these things, I usually hear from people who have had the same problem and are relieved that it wasn't just a programming error, as a referee or their thesis advisor had insisted. For some reason, however, we don't talk or write much about these puzzles.

It seems as if research in statistics has come to mean promoting new methods, as opposed to understanding methods, old or new. Obviously, it would be inaccurate and unfair to say that *nobody* tries to understand existing or new methods; counterexamples include the trace plots used to describe LASSO results or the separation plots used to explain the support-vector machine. But when a classmate and I tried to assemble a catalog of the mysteries and puzzles we've found as a rationale for more investment in understanding these new methods, we found it impossible to formulate anything that looked like a contemporary journal article. These days, a journal article needs to end in a triumph — "Behold! We have conquered this messy dataset/new class of models/previously intractable computing problem/[etc.]!" — and it is hard to make a mystery or puzzle look like a triumph. The triumphant narrative style is so embedded in today's conception of a journal article that even not-too-thoughtful extensions of linear model theory are cast as new methods. If you doubt this, try doing a literature search for outlier-detection methods for hierarchical or other random-effect models. With two exceptions that I know of (one that I wrote and one that I recently refereed), every such paper ends with a standard triumph in which the new method is shown to identify outliers, but none of these papers is based on a systematic understanding of models with random effects.

It's not hard to imagine why the literature looks this way. There are so many new models, where do you begin? With the present-day emphasis on generality, how can you do anything general enough to interest a good journal? The new methods are so complex! And so on. The solution to this quandary, it seems to me, is to stop trying to learn something about all or even a large fraction of the wonderful new methods developed in the last 30 years. They are too disparate and they are developing too quickly.

Having made that unsexy concession, there does seem to be a good place to start. Many of the new methods are now undergoing a process of unification analogous to the unifications that produced generalized linear models and, even earlier, the projection theory of linear models. The unification of models with random effects — so far — consists of a few competing syntaxes for expressing a large class of models and a method for fitting models expressed in each syntax. Parts I and II of this book emphasize one such syntax, mixed linear models using normal distributions, and some of the great variety of models, which I call richly parameterized models, that can be expressed this way and analyzed using conventional and Bayesian methods for mixed linear models. Examples of this unification include Robinson (1991) and Ruppert et al. (2003). This class of models is rich enough to be interesting and close enough to single-error-term linear models to allow many insights and methods to be borrowed or adapted.

A theory of richly parameterized linear models needs more than a syntax and a computing method. It needs to explain things that happen when these models are used to analyze data, to provide ways to detect problems, and when possible, to show how to mitigate or avoid those problems. Part III takes a step from the theory of ordinary linear models toward a theory of richly parameterized models by adapting ideas central to linear model theory. Part IV then takes a step beyond linear model theory by examining the information in the data about the mixed linear model's covariance matrices, which are the difference between ordinary and mixed linear models.

Parts III and IV are founded on two key premises. The first premise comes from linear model theory: Writing down a model and using it to analyze a dataset is equivalent to specifying a function from the data to the inferential or predictive summaries. However you rationalize or interpret that model, it is essential to understand, in a purely mechanical sense, the function from data to summaries that is implied by the model. Often when I present material from this book, people respond with things that are *non sequiturs* in these terms, for example: This model estimates a causal effect while this other model does not; or a Gaussian process probability model with such-and-such covariance function has realizations with such-and-such properties; or this posterior distribution or likelihood is by definition what the data have to say about the model's unknown parameters. I don't dispute such assertions but for the purposes of this book, they are irrelevant. The question asked here is: When I fit *this* model to *this* dataset, why do I get *this* result, and how much would the result change if I made *this* change to the data or model?

The second premise is that we must distinguish between the model *we choose* to analyze a given dataset and the process that *we imagine* produced the data or that, in rare cases, we know actually did produce the data. Our choice to use a model with random effects does not imply that those random effects correspond to any random mechanism out there in the world, and that fact has practical implications. These implications are a recurring theme of this book's first three parts and are summarized in Chapter 13.

Building on these two premises, Parts III and IV are organized around mysterious, inconvenient, or plainly wrong results that turned up in real problems. Most of these are from my collaborative work but some have come from colleagues, for example Michael Lavine's dynamic linear model puzzle (Chapters 6, 9, 12, and 17). Some of these puzzles are now understood to a greater or lesser extent, while others are barely understood at all. In that sense, Parts III and IV are not quite a catalog of unsolved problems — who knows how many puzzles are as yet unreported or undiscovered? — and their theory is grossly incomplete. I will apologize for that once, now. I hope Parts III and IV stimulate enough research so that the second edition of this book, if there is one, can report fewer mysteries and, yes, more triumphs.

Most statistics texts emphasize what we know. This book emphasizes what we don't know. Judging from reviews of my proposals, a lot of academics think mixed linear models are completely understood, when in fact they are still largely not understood. But this is good news: What a bounty of unsolved problems, and for a heavily used class of models! Graduate students, start your engines!

In emphasizing what we don't know, Parts III and IV consider open problems, which leads to a stylistic quandary. Traditionally, statistical theory follows a mathematical style emphasizing results that can be packaged as theorems. I have observed this tradition whenever possible, because you just can't do better than a relevant theorem. Unfortunately, I could not always produce theorem-like material in a reasonable amount of time. In situations like this, statistical theorists usually present nothing at all or present something they *can* package as a theorem, most often some kind of large-sample result. That seemed unproductive, so when I haven't been able to produce relevant theorem-like results, I have instead followed a style used by my scientific (as opposed to statistical) colleagues. They work by posing hypotheses and gathering a variety of evidence, often indirect, so that their hypotheses are either refuted or accumulate credibility while becoming more refined. Obviously accumulated credibility can't replace the iron-clad certainty of a theorem when a relevant theorem can be proven, but the new methods of the last three decades are so complex that it may never be possible to prove relevant theorems about them. We can, however, make progress by approaching our black-box methods in the same way our scientific colleagues approach nature's black boxes, by prying them open gradually and indirectly if necessary. Chapters 11, 12, and 17 are examples of this style, with Chapter 11 refuting a hypothesis and Chapters 12 and 17 developing some hypotheses and producing a first increment of credibility for each.

Along with a sometimes non-mathematical style of inquiry, I've written in a relatively informal narrative style. I've done so because it's friendlier in two senses: It's easier to understand on a first reading and it doesn't hide my opinions and ignorance behind the passive voice and calculated omissions. Students and other readers *should* find fault with the current state of this field, including the things I've contributed to it, and I want them to see those faults and decide they can do better. I will be delighted if this book attracts the interest of young people with better math and computing skills than I have, who can change this area of study from the backwater it is into the thriving area it can and should be.

Some Guidance about Using This Book

The object of Part I is to present a survey of essentials and a particular point of view about them. The object of Parts II, III, and IV is to present the beginnings of a theory of richly parameterized linear models. This book is not intended to be a magisterial overview of everything known about mixed linear models. It is rather intended to present a point of view about what we do and do not understand about mixed linear models and to identify research opportunities. Overviews of mixed linear models include Searle et al. (1992); Ruppert et al. (2003), which focuses on penalized splines represented as mixed linear models; Verbeke & Molenberghs (1997), which focuses on SAS's MIXED procedure; Diggle et al. (1994), which focuses on longitudinal models; Snijders & Bosker (2012), a thorough treatment of hierarchical (multi-level) models obviously based on a lot of experience fitting them and explaining the fits; and Fahrmeir & Tutz (2001), which catalogs models with exponential-family error distributions and linearity in the mean structure.

The hazard of writing a book like this is that I have to write short chapters about sub-fields of statistics with huge literatures. Even D.R. Cox might not be able to master all those sub-fields; I certainly haven't. Academics tend to be territorial and to view the world through a microscope, so whatever I write is guaranteed to offend specialists in each sub-field even if I say nothing that is factually incorrect. Also, the literature and folklore of mixed linear models are gigantic and I know less than

I probably should about them. Therefore I ask your indulgence: I have tried to be nice to everybody and in return, if I've said something factually incorrect or wrong-headed, please tell me and provide detailed citations. If I decide you're right, I'll post a suitable piece on the book's web site with credit to you and replace the relevant passages in the next edition, if there is one.

I've tried simultaneously to give results about both Bayesian and conventional (non-Bayesian) analyses which, these days, mostly revolve around the restricted or residual likelihood. I've done this because Leo Breiman (2001) was right: The two approaches really aren't much different in practice, at least in this area. Chapter 1 is my argument for that claim. I had the good fortune to study in a department where I could become fluent in both languages but most people aren't so fortunate and thus might find it difficult to switch back and forth between Bayesian and conventional language. I've tried to make this as clear as I can and I apologize for any failures of clarity.

I wrote this book from classroom overheads for a one-semester course that I teach for advanced PhD students in the Division of Biostatistics at the University of Minnesota. I use *Semiparametric Regression* by Ruppert, Wand, & Carroll (2003) as a textbook for that course and Parts I and II of the present book refer to it frequently. *Semiparametric Regression* is simply lovely. Among statistical books with hard technical content, it is the friendliest I've ever read and I only disagree with three or four things in the whole book. I recommend it without reservation. If the present book is written half as well, I will be happy.

Each of the present book's chapters ends with exercises, which are of two types. The first type is standard results that PhD students should be able to derive, which are intended to provide practice with the mostly algebraic methods used in this book. Most chapters also include exercises that are, as far as I know, open research questions. Often these include suggestions about where to start, but of course you should feel free to ignore my suggestions.

The book's web site includes datasets analyzed as examples, when I could get permission to include them. I will be happy if you find these datasets useful and horrified if any of them ends up being pawed over eternally like the stack-loss data or the Scottish lip-cancer data. Publish your own datasets! The world will be richer for it.

When I have used one of my published analyses as an example, I have presented it the way it was published. I figured it would be both dishonest and hazardous to make myself look smarter than I actually was and hope nobody checked the original papers. This also gave me an incentive to be nicer to other researchers than I might otherwise be. In each such case I point out what I now believe is wrong with the analysis I published and when it seems worth the effort and space I give a better analysis. If you identify blunders I haven't mentioned and I agree they are blunders, I will post your attempts to alleviate my ignorance on the book's web site.

Finally, Part I refers to SAS frequently because I am in a biostatistics department and even though cognoscenti are obligated to sneer at SAS, we teach it to our students (along with R and WinBUGS) because it helps them find jobs. I don't mean to single out SAS for criticism but it *is* the Microsoft Office of statistical software and that makes it a good example for many purposes.

May 2013