Although the book manifests some important omissions, it contains a good deal of material that will be useful to statisticians and social scientists interested in learning more about causal inference. Its strengths include its clear exposition, scope, inclusion of econometrics material, and interesting examples. The book fills an important gap in the field, since prior to its publication there was no book on causal inference that even made an attempt at being comprehensive and/or representative of the literature. Morgan and Winship have rendered a useful service to the field in making what is a rapidly expanding literature accessible to a wider audience. I would highly recommend the book as an introduction to causal inference.

Tyler J. VanderWeele
Harvard University

REFERENCES


Should you consider reading this book? To decide, consider the following cluster of assertions: In statistics, “significant” is not the same as important. Statistical analysis cannot be reduced to fixed-level testing. For decision problems, well-chosen loss functions can be indispensable. Many complex modeling challenges lend themselves to a Bayesian approach. Almost always, effect sizes and interval estimates matter.

If you agree, you will probably find little that is new in The Cult of Statistical Significance except for the questionable entertainment value of the evangelical excesses, which begin with the subtitle How the Standard Error Costs Us Jobs, Justice, and Lives. You can just imagine the tabloid headline: “ATTACK OF THE KILLER SEs.” The Cult... comes close: Fisher is “our flawed villain” (p. xv) and “the wasp.” And, in more detail: “Fisher-significance is by itself about precisely nothing. ... Statistical significance ... has collapsed the scientific world into a Borel space, p(0, 1.0)—a procedure, by the way, that the mathematical statistician Emile Borel (1871–1956) himself emphatically rejected” (p. 9). “...today’s statistical experts do not estimate or consider the loss function or Type II error at all. ...without a loss function a test of statistical significance is meaningless, no better than a table of random numbers” (p. 8).

In my first paragraph, I have given you the intellectual substance of the book. In the second paragraph I have given you the flavor of the authors’ analytical and rhetorical style. Based on substance and style, you know enough at this point to decide whether the book might interest you. If your only concern is to decide whether to take a look at this book, you can stop reading.

If you can stop reading, why did I keep writing? What else can I offer readers of JASA that might be worth your time and the journal’s space? Answer: I hope to convince you that this book poses a double threat to our profession, and that we all have a responsibility to inform ourselves so that we can respond if colleagues from the sciences ask us about the book. I confess that I found it deeply disturbing. Off and on over the last several weeks, I have thought hard about the reasons for my reaction, and what, if anything, to write in response. I’ve dumped two draft reviews into the trash. Here is a five-point distillation of my current thinking:

1. The issues addressed by the book are of vital importance, not just to our own profession, but more broadly to all of the natural and social sciences, and so, indirectly, to everyone who benefits from the advance of knowledge in these areas.
2. The authors of the book are well-known senior professionals in an area—economics—that is closely tied to our own, and their main thesis, as summarized in my first paragraph, is one that deserves broad acceptance.
3. Sadly, the book they have written is intellectually sloppy: the exposition of statistical issues is mushy, the rhetoric is partisan and overblown, the meta-analysis is unscientific, and the history careless at best.
4. These shortcomings give rise to the first threat: Because the issues are both deep and consequential, I worry that the authors’ tabloid style may lead serious scientists to dismiss their thesis. (The authors apparently worry as well, with a defensive expectation—p. 31—that readers “merely get angry at our style.”)
5. There is a second, graver threat, that scientists who should be making statistical thinking a central part of planning their experiments and understanding their results, but who do not understand why statistics matters, and who feel defensive about their lack of conceptual understanding, will embrace this book as a license to write off statistics as worse than useless. In this sense The Cult... could turn out to be deeply destructive, not only to our own profession, but to all subject areas whose advance depends on an empirical approach to understanding. Because the issues are so important we statisticians should take ownership of the issues.

In what follows, I address each of these five numbered points in sequence, then conclude with a brief for optimism.

(1) The issues are of vital importance. In one sense the issues, as summarized in my first paragraph, are not so much vital as dead—as in a horse so moribund that only a zealot would muster the effort for a book-length flogging. Apparently, however, these truths are not self-evident, and, according to Ziliak and McCloskey, many reputable journals in the sciences publish articles by many reputable users of statistics whose using is narrowly limited to yes/no questions of significance. For reasons I set out below, I remain skeptical of the authors’ assertions that current statistical practice is still dominated by such an antiquated, narrow and exclusive reliance on whether $p < 0.05$. However, my skepticism is based not on data, but rather on an instinctive distrust of any claim of fact when the assertions come wrapped in such highly partisan rhetoric. Extreme vehemence invites suspicion. For sound conclusions, we need good data on bad practice.

(2) The authors are well-known senior professionals. First author Stephen Ziliak is Professor of Economics at Roosevelt University; he has written two books. Deirdre McCloskey is Distinguished Professor of Economics, History and English at the University of Illinois at Chicago. She has held Guggenheim and National Humanities Fellowships, and has written 20 books. The Cult... is an expansion of articles published in various journals, including the prestigious American Economic Review.

(3) The book is intellectually sloppy: (a) the exposition of statistical concepts is mushy, (b) the rhetoric is partisan and overblown, (c) the meta-analysis is unscientific, and (d) the history is at times careless at best. Here we come to the nub of my discontent. I applaud the authors’ dedication to sound statistical practice. I agree with their assertion that sound practice must involve thoughtfulness, and interval estimates matter.

(4) These shortcomings give rise to the first threat: Because the issues are both deep and consequential, I worry that the authors’ tabloid style may lead serious scientists to dismiss their thesis. (The authors apparently worry as well, with a defensive expectation—p. 31—that readers “merely get angry at our style.”)

(5) There is a second, graver threat, that scientists who should be making statistical thinking a central part of planning their experiments and understanding their results, but who do not understand why statistics matters, and who feel defensive about their lack of conceptual understanding, will embrace this book as a license to write off statistics as worse than useless. In this sense The Cult... could turn out to be deeply destructive, not only to our own profession, but to all subject areas whose advance depends on an empirical approach to understanding. Because the issues are so important we statisticians should take ownership of the issues.
(a) The exposition of statistical concepts is mushy. Readers of JASA no doubt recognize a variety of ways one can misuse significance tests: for applied decision problems, one should be thoughtful about choosing a loss function; it would be wrong to rely instead on a test of significance. In a different context, estimation rather than testing might be the appropriate focus, and a Bayesian approach may be indicated. Here, too, it would be wrong to use a significance test. In yet other contexts, a significance test may indeed be appropriate, and some discussion of the power of the test against alternatives may be relevant.

Whether an analysis should involve loss functions or Bayesian methods or a discussion of power will depend on the context and the goal of the analysis. No one would argue that significance testing is the method of choice for all statistical applications. [See, e.g., the thoughtful distinction that Arthur Dempster makes (1971) between "IS questions—IS there an effect?" questions that focus on model selection and lend themselves to Fisherian testing, and "IT questions—how big is IT?", questions that take the model as given, seek to estimate likely parameter sizes, and lend themselves to a Bayesian approach.]

 Nevertheless, the authors do not try to distinguish among these different situations. Instead they write as though the use of \(p\)-values without as loss function is prima facie evidence of poor practice, as is the failure to mention power or to use a prior distribution.

(b) The rhetoric is overblown. The second paragraph of this review offered a small sample. Here are a just a few more instances, harvested from an embarrassment of riches.

- \(R^2\), \(t\)-statistic, \(p\)-value, \(F\)-test, and all the more sophisticated versions of them... are misleading at best” (p. xv).
- “The average article in the leading journals of the statistical sciences claims that size doesn’t matter” (p. 41).
- “...almost all the teachers of econometrics claim that statistical significance is the same thing as scientific significance” (p. 107).
- “We were alarmed to find a scholar and statistician of the quality of Bradley Efron saying that ‘we could badly use a new Fisher to put our world in order’...” Another Wasp is precisely what we do not need” (p. 247).

The rhetorical excesses might be dismissed as mere froth but for the book’s claim to be a serious examination of the use and misuse of statistics. The over-reaching, unfortunately, is not limited to mere rhetoric. Core chapters on meta-analysis and history are also undermined by overreaching with regard to substance.

(c) The meta-analysis is unscientific. Chapters 5–9 are based on a study of 20 years’ worth of articles published in the American Economic Review. Every full-length article published between 1980 and 1999 was scored using a 19-item yes/no questionnaire designed by the authors to rate the use of statistics on a scale from 0 (worst) to 19 (best). For the articles published in the 1990s, they find (pp. 90–91):

| Exemplary | 15–19 | 12 articles |
| Good      | 12–14 | 26 |
| Fair      | 9–11  | 42 |
| Poor      | 6–8   | 69 |
| Very poor | <6    | 37 |

I have no reason to doubt the authors’ main point, that all too often, the use of statistics—mainly regression—in these articles may not have been particularly thoughtful, and could have been improved by more attention to descriptive statistics and to the possible meaning of the regression coefficients in the context of the application at hand. At the same time, however, I do not find the authors’ method particularly compelling. The 19 questions, they claim, “are posed so that a yes answer means that the practice is good, ‘good’ by the standards of any serious statistician” (p. 66). Apparently, the validity and reliability of the scale are taken to be self-evident. “The criteria are not controversial.” And yet the first question asks about sample size, and awards a point if the sample size is small. (The dubious logic: Because a very large sample can pretty much guarantee statistical significance, a \(p\)-value is meaningful only if the sample size is modest or small.)

Other questions appear reasonable: Are the units and descriptive statistics presented for all variables that appear in a regression? Are the proper null hypotheses specified? Does the article refrain from using statistical significance as the criterion of scientific importance?

Although some of the individual questions may seem reasonable, there are some striking omissions: There is no mention of regression diagnostics, or exploratory analysis, or graphics, or interval estimates. Moreover, there is a great deal of overlap among the 19 questions. Five questions deal with reporting the size of regression coefficients: does the article refrain from ranking variables by \(R^2\)? ... refrain from noting sign but not size of coefficients? ... discuss size at all? ... discuss size in relation to applied context? ... choose variables based only on \(p\)-values? Another five questions deal with reporting significance: is the first use of significance merely one of multiple criteria? Do authors refrain from using statistical significance as a criterion for scientific importance? Is statistical significance portrayed as decisive? In the concluding section, is statistical significance separated from importance? Is the use of the word “significance” unambiguous?

Evaluating any given article calls on the rater to make many subjective judgments: Which are the “proper” null hypotheses? What constitutes an “ambiguous” use of the word “significance”? How does one judge whether coefficients have been interpreted “carefully”? or judge “when a test of significance is not relevant”? The ratings were done by the authors themselves, rather than by blind judges, and there is no attempt to measure inter-rater reliability.

(d) The history is careless at best. (Here I am relying on help from Stephen Stigler and Bradley Efron.) The book’s narrative climax is a three-chapter recounting of the relationship between Fisher (“the Wasp”) and Gosset (“the Bee”), culminating in Chapter 22’s fable “How the Wasp Stung the Bee and Took Over Some Sciences.” Here Ziliak and McCloskey tell a story of how—according to them—Fisher claimed credit for himself for the table of values of \(t\) in his 1925 article “Expansion of ‘Student’s’ Integral in Powers of \(n^{-1}\)” published in Metron. They write (p. 229) “...the sting in Metron is that the three tables of \(t\) invented and then calculated nearly entirely by Gosset, are attached to Fisher’s... article... as if Fisher was the inventor and calculator.” Three pages later (p. 232), referring to the way statisticians have often attached Fisher’s name to the \(t\)-test, they write “Fisher did not ever, it appears, acknowledge the mass confusion he had caused. It would anyway have been uncharacteristic of him to have done so. He would have had to admit straightforward scientific fraud.”

Straightforward scientific fraud? Here is an account kindly provided by Stephen Stigler, whom readers will recognize as one of our profession’s most distinguished historians:

In 1922–25 Gosset and Fisher worked together to compute much more detailed versions of the earlier \(t\)-tables: more values of \(t\), more \(df\), more decimal places. This seems to have been entirely cooperative-based on the one side of the correspondence that survives (Gosset’s letters to Fisher, kept by Fisher). Fisher developed new expansions (in powers of \(1/n\)) and his assistant Miss Mackenzie computed the coefficients for the expansions at Rothamsted, and Gosset did the grunt work of calculating and checking via differences etc. Or so I surmise from the correspondence. Karl Pearson would (I similarly surmise) have been willing to publish but he was steadfast in keeping copyright so the Tables books he published would not lose too much money. Fisher was leery of this and sent them to Gini’s journal Metron. Three papers appeared there in 1925—Fisher on the relationship of the distributions (\(t\), chi-square, etc.), Student with the table, and Fisher with the expansion used (Fisher 1925b). I did not go back to Metron yet but this is from their collected works (v. 2 for Fisher). Also in 1925 Fisher published the 1st edition of Statistical Methods for Research Workers (SMRW: Fisher 1925a), which included a \(t\)-table as Table IV and (apparently—I have a 1st ed and don’t see an explicit copyright) copyrighted the whole book under his own name. Student is cited generously with regard to the test but not mentioned with regard to the table. That is entirely appropriate. The table in SMRW is entirely different, and as far as I can tell recomputed by Fisher. The Metron tables are much smaller and gives \(t\) for various \(df\) and \(n\). The SMRW table is much smaller and gives \(t\) for various \(df\) and \(p\). Metron tables the distribution function; SMRW tables the inverse, the percent points, and it is this form that has caught on. Now conceivably Fisher could have got some of his (2 sided) \(p\) values from interpolation in the 1 sided Metron table, but I just made a spot check and I can tell you it was not linear interpolation, and indeed some values are impossible to get from the Metron table by any form of interpolation. Fisher was a far superior applied mathematician to Gosset and with his assistant at Rothamsted computed all the other SMRW tables so why not this one? And anyway, copyright applies to the form of expression (totally new here), not the underlying distribution. Fisher had had Gosset read proofs for SMRW, and there is no sign of discontent on this in their correspondence.

Not to put too fine a point on it: the book by Ziliak and McCloskey builds, over 20 chapters, to the culminating accusation that Fisher committed “straightforward scientific fraud” by stealing credit for Gosset’s work. Yet the published...
documents and surviving correspondence indicate that Fisher and Gosset had worked collaboratively, that the tables that Ziliak and McCloskey accuse Fisher of stealing were substantively different from those created by Gosset, and were in fact based on Fisher’s own work. (For more detailed historical accounts and commentary, see Stigler 2008 and Zabell 2008. For a review of the book written for scientists, see Porter 2008.)

Sadly, such a major misconception of the historical record, combined with the other features of sloppy scholarship set out in (3), may lead serious readers to throw out the baby—to dismiss the main thesis (4 below), or even worse, may lead gullible readers to swallow the bathwater—to accept the main thesis (5 below).

(4) Threat 1: Serious users of statistics throw out the baby. When I think of this threat, I have in mind a composite colleague, a mixture of people, all of whom I like and admire, some of whom used to advise their economics majors to avoid my applied regression course because “he wants you to look at the data.” The sanctioned approach was: “Model comes from theory. Fit the model, determine the variables that register as significant, and look at the sign of the coefficient: is the sign what theory predicts?” This practice is exactly what Ziliak and McCloskey rail against. (If only they had been advising my college’s economics majors, my applied regression enrollments would have been much larger.) But: for those readers of The Cult who are like my composite colleague, I worry that the sloppy scholarship of The Cult will make it all too easy for them to dismiss the much-needed challenge to rethink their narrow paradigm for applied regression.

Bottom line: The Cult makes it all too easy for the thoughtful reader to dismiss as junk what matters most.

(5) Threat 2: Gullible readers swallow the bathwater. For this threat, I have in mind a different composite colleague, one who, though also likeable, and in many ways both intelligent and committed to sound science, is fundamentally overwhelmed when it comes to statistics. Most likely due to a history of bad teaching, this reader’s understanding of statistics remains mired in the swamp of computation. A lifetime of unrewarded struggle has primed this reader’s psyche to clutch at reasons to dump the whole statistical enterprise as worse than worthless. For such a reader, The Cult is a glorious vindication: Ziliak and McCloskey have written a book that supports what you’ve known all along—statistics is just painful; it’s downright evil, in that it actually retards scientific progress.

Moreover, although this hypothesis of statistics-as-evil may be presented as the original thesis of Ziliak and McCloskey, now that they have put their necks out, they summon an honor roll of predecessors who have, anachronistically, lined up on their side and “agreed with our main point”: “Zellner, Kruskal, Rothman, Leamer, Hirschliefer” (p. xix), “edgeworth, Gosset, Egon Pearson, Jeffreys, Borel, Neyman, Wald, Woffowitz, Yule, Deming, Yates, Savage, de Finetti, Hirschliefer, Good, Lindley, Feynman, Lehmann, DeGroot, Bernardo, Chernoff, Raiffa, Arrow, Blackwell, Friedman, Morris, Hjort, Lo, Cox, Savage, Tukey, Kruskal, Mandelbrot, Wallis, Roberts, Granger, Press, Moore, Berger, Freedman, Rothman, Leamer, and Zellner.”

Bottom line: The Cult makes it all too easy for the statistically intimidated reader to accept as gospel what matters least.

(6) Conclusion: Grounds for optimism. If you’re still with me, I hope you have in fact succeeded in raising your blood pressure. What action should we take? My hope is that vigilance alone will suffice. According to Thomas Kuhn’s Structure of Scientific Revolutions (Kuhn 1962), rethinking the basics of science comes not so much from changing the minds of those who currently steer the enterprise as from waiting for them get old enough or tired enough or dead enough to cede the wheel to younger drivers who grew up with a different sense of direction. On that hypothesis, the future use of statistics in the sciences will be largely shaped by the current teaching of statistics in our high schools and colleges. Hence my optimistic outlook. Forty years ago, when I was an undergraduate, the teaching of statistics and “look at your data” lived in complementary subspaces. A mere 10 years later, a seismic shift was underway. Within statistics proper, “Look at your data” would soon span the whole space. The orthogonal complement—“Real data? What real data?”—was contracting to zero. I am optimistic in thinking that over the next two decades, the current approach to data-based teaching will come to dominate the practice of statistics across our domain.

George COBB
Mount Holyoke College

REFERENCES


Demographic Forecasting.


With their new book, Girosi and King contribute a new Bayesian methodology of forecasting mortality when time series are noisy and sparse. The method is strongly evocative of seemingly unrelated regression (Zellner 1962); a somewhat less fortunate parallel is that the book’s title is almost seemingly unrelated to the method’s much broader applicability. This is not to understate the authors’ valuable contributions to the particular field of mortality forecasting, as evidenced by extensive and repeated application of the methods. While Girosi and King have sidestepped altogether the arguably more thorny issues of what to do about the other two components of demographic change, namely fertility and immigration, their insights into mortality forecasting are certainly deep and innovative enough to make this book a valuable and welcome addition to the subfield. But it is not until the last section of the introductory chapter that we catch a glimpse of the wider relevance of the methods, during a discussion of their applicability to, and perhaps their partial genesis in, a seemingly unrelated field: comparative political science.

The acknowledgements section makes clear the authors’ specific motivation to improve mortality forecasts for developing countries, even if it almost sounds like they did it on a dare, and their contribution is a great success in this regard. Prudent demographers will forgive the occasional gibes about their getting the order of applying priors exactly backward; Girosi and King make the excellent point that knowledge about the smoothness of mortality rates across ages, across borders, and even across time ought to be applied before the model is estimated, not only as a check on the results. Bayesian priors can accomplish this; another method forecasters use is to smooth the data via splines and then account for that when fitting (Currie, Durban, and Eilers 2004). A second major insight is that because we expect this smoothness in mortality, which is our dependent variable, standard Bayesian methods that incorporate smoothness priors about model coefficients need some tweaking. The authors show how to do this theoretically and operationally, via Markov Chain Monte Carlo and some faster alternatives, in Parts II and III, respectively. The broadness of the method’s appeal really hinges on a third insight that is related to the second: when mortality data are sparse, covariates like income or smoking are useful, but their marginal effects may vary considerably across countries. This is the hook for comparative political science, a field in which pooling across countries is apparently a mortal sin for those who are not armed with the latest in Bayesian techniques. Everyone, please do not tell political scientists what we economists are up to on a daily basis.

Part IV reveals the model in action, going toe-to-toe exclusively with the widely used model of Lee and Carter (1992), which the authors described in their review of the literature in Part I. They would be the first to admit that
Lee–Carter was never designed to work on poor-quality data outside the United States or G-7, or on cause-specific mortality; they show here that their methods can generate results that indeed look much more “reasonable,” like the Bayesian priors encapsulate, than those of Lee–Carter in these cases. The rough linearity of the Gompertz slope in log mortality is maintained over time, as any wild fluctuations that may be in the data melt away. There is debate about the wisdom of producing cause-specific forecasts given problems with bias at least over long periods (Wilmoth 1995), but it is clear that there is great demand for them in public health circles. A recent review by Booth (2006) is useful in assessing the pitfalls. The penultimate chapter assesses model performance by fitting a data prior to 1990 and testing on data since then. The new method improves considerably on Lee–Carter exactly where it was intended, among developing countries and in cause-specific mortality. Although Li, Lee, and Tuljapurkar (2004) have showed Lee–Carter can work when time series are short, the great advantage of Girosi–King is that it looks like it works when cross-sections are noisy.

Naturally, the book leaves some questions unanswered. We do not know how Girosi–King forecasts of future life expectancy compare to official Social Security forecasts, for example, although one suspects that like Lee–Carter projections, which are also extrapolative, they are comparatively optimistic. While the debate over optimism and pessimism seems unlikely to be unequivocally resolved until it is conducted ex post, it would still be useful to know where this method stands. It is not clear how smoothing priors may or may not change the predicted future of the global demographic transition and population aging, although this is surely a cheap shot here given the relatively large importance of the fertility transition. More practically, we do not know how much the Girosi–King method will help guide policy, because it is not clear how much it will help forecasters in government agencies, who are typically strapped for time and face impediments to incorporating innovative forecasting techniques. The present value of Demographic Forecasting seems mostly to derive from its very insightful approach to a specific problem that is isomorphic to many others in applied inference, although the work is clearly of significant interest to specialists forecasting mortality in developing countries.

Ryan D. Edwards
Queens College and the Graduate Center,
City University of New York and NBER

REFERENCES

Introductory Lectures on Fluctuations of Lévy Processes With Applications.


Introductory Lectures on Fluctuations of Lévy Processes With Applications is a textbook that systematically presents the path properties of Lévy processes and their applications. This book is written in a pedagogical style with many exercises and figures. According to the author, the intended audience for this book is PhD students with basic knowledge of real and complex analysis, Lp spaces, measure theory, Markov processes, Poisson processes, Brownian motion, and continuous parameter martingales. However, because of the lucid presentation, it should also be useful to researchers with the same mathematical background and interested in applications of fluctuation theory of Lévy processes.

This book consists of 10 chapters with 22 figures and approximately 80 exercises. The first chapter contains a brief introduction to Lévy Processes, their relationship with the infinitely divisible distributions and various applied probability models based on them. Chapter 2 focuses on the Lévy–Itô decomposition and describes the structure of a general Lévy process in terms of three independent Lévy processes with different path properties. In Chapter 3, the author examines the Strong Markov Property, duality, moments, and exponential change of measure of Lévy processes. The fourth and fifth chapters give applications of Lévy processes to queuing theory and renewal theory, respectively. The sixth chapter provides an account of the theory of excursions of a Lévy process from its maximum and in particular the Wiener–Hopf factorization, which is used in Chapter 7 to characterize the behavior of Lévy processes at first passage over a fixed level. In Chapter 8, the author specializes on spectrally negative Lévy processes and their exit problems. Chapters 9 and 10 describe the applications of Lévy processes to optimal stopping problems and continuous state branching processes, respectively. Towards the end, the author mentions several related areas and a few key references in each of them. This “épilogue” is extremely useful to readers who would like to pick up a research problem on this topic. Overall, this book is mathematically rigorous and yet gives intuition for underlying concepts and results.

Although it may be used as a textbook for an advanced graduate level course on the fluctuation theory of Lévy processes, none of the exercises given in this book can be assigned in a homework or a take-home examination because their solutions are given in details in the book itself. However, it is useful in a seminar course or a reading course with a smaller group of students, who are expected to learn and present the subject material as well as the solutions to the exercises. Though some of the exercises are rather long and time consuming, most of them do help the reader to understand the underlying concepts better. As mentioned by the author, advanced readers will find the books of Bertoin (1996) and Sato (1999) more sophisticated than this book.

In summary, Introductory Lectures on Fluctuations of Lévy Processes With Applications is a useful book for graduate students and other researchers wishing to become better acquainted with the path properties of Lévy processes, as well as to instructors who are willing to offer a seminar (or reading) course with a view to training a group of PhD students for research on this topic. The lucid and pedagogical style of writing and a careful balance of mathematical rigor and informal discussions make this book an extremely valuable addition to the library of anyone involved either in the theory of Lévy processes or its applications to storage models, risk theory, and mathematical finance.

Parthanil Roy
Michigan State University

REFERENCES

Linear and Generalized Linear Mixed Models and Their Applications.


This book is a detailed and comprehensive account of the mathematical structure of generalized linear mixed models (GLMMs) and their fitting to data by maximum likelihood (ML) and other methods. The book, of 257 + xiv pages, is set out in four main parts: two on Linear Mixed Models (LMMs) and two on GLMMs. Part I of LMMs discusses model formulation and parameter estimation, by maximum likelihood and other methods, and gives two examples. Part II covers hypothesis testing, confidence interval estimation, prediction, model checking and selection, and Bayesian inference, and discusses two examples. Part I of GLMMs discusses model formulation in the exponential family, the difficulties of the likelihood function, several methods of approximate inference, and prediction of random effects. Part II discusses likelihood-based inference, generalized estimating equations and model selection, with several complex examples. The last parts have extensive technical notes, many of which were derived by the author and his co-workers. Some of them are at a high level of mathematical difficulty; the author does not shirk complex derivations where needed.
There are three appendices on notation, matrix algebra, and some statistical results on quadratic forms and the exponential family. There are more than 200 references, many of them recent.

The author’s approach to inference is eclectic: ML, non-ML iterative weighted least squares, Bayesian, bootstrapping, and jackknifing all appear, and the level of analytic detail about these approaches is impressive. The book is a valuable reference for many difficult analytic results, and would be suitable for a course featuring a theoretical treatment of GLMMs. Each chapter has a large number of exercises validating or extending the results in the text, a very useful feature for instructors.

However, the instructor would need to supplement the book with more detail for data analyses, and particularly for Bayesian analyses of these models. The author gives a useful list of recently published books on mixed effect models. For a wider coverage of applications the books by McCulloch, Searle, and Neuhaus (2008), Verbeke and Molenberghs (2000), Molenberghs and Verbeke (2005), and Pinheiro and Bates (2000) would be useful. The author notes that most of these books do not deal with non-Gaussian random effect models; references that describe maximum likelihood analysis of some of these models can be found in Aitkin et al. (2009). A very wide coverage of all of the models in the book, and others involving finite mixtures, latent variables and non-Gaussian random effects, can be found in the comprehensive book by Skrondal and Rabe-Hesketh (2004).

The instructor would also need to alert to several errors in the author’s analysis of GLMMs. The first is in the computation of the likelihood with random effects. The author states (p. 126) that, for the salamander data with two crossed random effects with 40 levels of each, the likelihood is the product of 1600 terms all of which are small, so the likelihood computation would give zero. This is incorrect. The computations for the EM algorithm are carried out for the complete data log-likelihood, which avoids the underflow problem, which can also be circumvented in the computation of the observed data log-likelihood.

A more serious omission is any discussion of nonparametric maximum likelihood, in which the random effect distribution is estimated nonparametrically, as a discrete distribution on a finite number of mass points. This is relatively straightforward, as the discrete quadrature is very similar to Gaussian quadrature (GQ), and starting values for an EM algorithm for the nonparametric case can be taken from the GQ analysis (Aitkin 1999; Aitkin, Francis, and Hinde 2005, chapter 9; Aitkin et al. 2009; Skrondal and Rabe-Hesketh 2004). This omission is unfortunate as the impression is left that other approximate non-ML methods are necessary for (non-Bayesian) analysis when the random effect distribution is unknown or nonnormal.

A third issue is the precision of empirical best linear unbiased predictors (EBLUPs), or empirical Bayes (EB) predictors, of the random effects. A detailed discussion is given (pp. 74–80) of the need for better estimates of the precisions of the usual EB estimates (posterior means) of the random effects. This is a point of major practical importance in many fields, where “league tables” of posterior means are widely used, incorrectly, for ranking. Fully Bayesian procedures are essential for this, as Markov chain Monte Carlo (MCMC) draws from the posteriors account fully for the uncertainty in the estimates of the variance components. It is therefore disappointing to find that the fairly brief discussion of Bayesian procedures for the normal variance component model (pp. 99–102) does not mention this possibility, and instead points to the agreement between EB estimates from Bayes analyses with known variance components and EBLUPs from restricted ML (REML) analyses. This is more surprising because the discussion of Bayesian methods for GLMMs in Part II has a detailed description (pp. 163–183) of both Monte Carlo EM algorithms and MCMC methods for exponential family models.

The book is less successful in the analysis and discussion of practical examples. On p. 37 the author gives a two-level REML analysis, using SAS PROC MIXED, of lamb birth weights with sires as the random effects and two categorical fixed effects, line (5 levels) and age (3 levels). He notes (p. 36) that it is important to fit the model without an intercept, because otherwise line and age effects are not all identifiable. He finds (Table 1.3) that the two age effects are nonsignificant, whereas all five line effects are significantly different from zero, and concludes that the average birthweight of the lambs differs from line to line.

This conclusion does not follow from the table estimates. It is clear that without a model intercept, SAS has aliased the third age category to zero, while the line effects are not differences relative to a reference category, but are the line means. This simple error could easily confuse a student unfamiliar with aliasing.

A second analysis is given (p. 37) of the lamb data with the MIVQUE0 option instead of REML. The author notes that the sire variance components are quite different in the two analyses, and a third analysis with analysis of variance estimates of the variance components (Henderson’s Method III) gives different estimates again. No comment is made on these differences, so the student is left unclear which (if any) is best, or whether these differences matter.

In the GLMM examples, one is of a dataset from a Biometrics paper on fetal mortality in mouse litters. The paper authors used a beta-binomial model to fit the overdispersed mortality counts (and a mixture model for outliers, not discussed in this book). They noted that the beta-binomial model fit to this data set is suspect because of a high proportion of outliers. The author fits an overdispersion model with normal random effects on the logit scale, and gives estimates of the mean and standard deviation (SD) of the normal distribution, obtained by a robust asymptotically optimal method.

The purpose of the analysis is unclear. Are the ML estimates of the mean and SD invalidated by the outliers? If not, are the mortality counts modeled better by the logistic/normal model than by the binomial/beta model? The answers to these questions would be interesting and useful, but would require at least a maximized likelihood from both models to assess. A robust analysis does not provide this; the student is left uncertain whether the robust analysis is better, or needed, and whether robustness comes at some efficiency cost.

These deficiencies are perhaps minor compared to the substantial theoretical contribution of the book. I hope that a second edition will improve the treatment of applications, and give more space to efficient Bayesian computation.

**Murray Aitkin**

**University of Melbourne**

**REFERENCES**


**Matrix Methods in Data Mining and Pattern Recognition.**


This book is a gentle introduction to linear algebra techniques which are widely used in data mining applications. It is mainly directed at undergraduate students, especially those with some background in basic numerical linear algebra. The presentation is nevertheless pleasantly self-contained, since the most basic concepts in scientific computing and in linear algebra are reviewed in a concise, clear, and effective way.

The book is of interest to a wide audience, including students in applied mathematics and statistics interested in learning about numerical techniques and their applications to data mining, as well as students already knee deep in applications, for example, in engineering, computational biology, and computer science. Students who wish to understand existing techniques, both in terms of their foundations and their applicability, will find this text valuable. Many departments have introduced or are planning to introduce interdisciplinary courses for this audience, and this book would be a good textbook. It provides sufficient background and could be easily integrated with research papers and projects.

This is an important achievement for a book directed to an undergraduate audience, and especially crucial in an interdisciplinary and rapidly evolving field such as data mining.