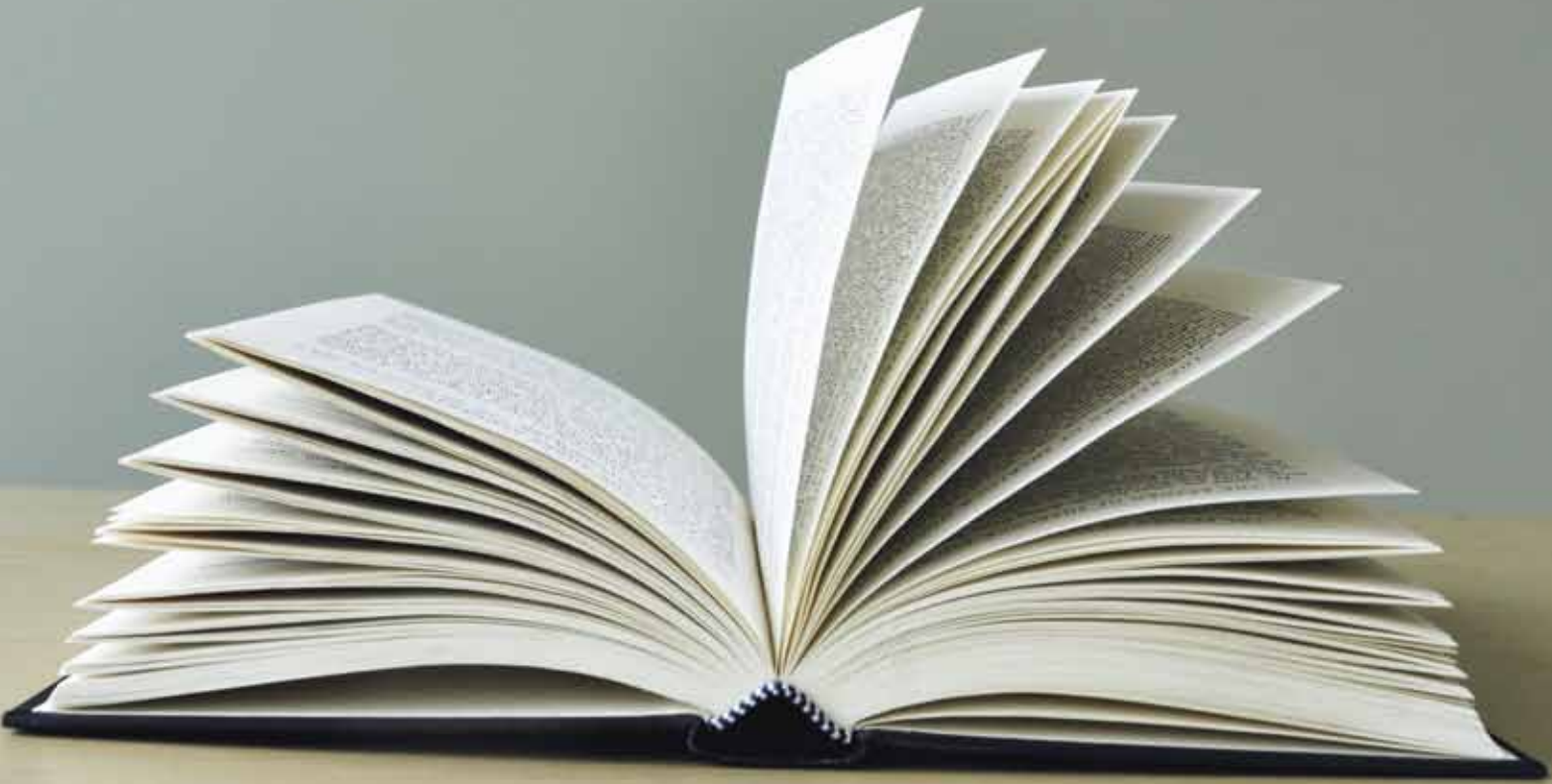


Special Issue on
George Casella's Books



On June 17, my dear friend George Casella passed away after a long illness that he fought with his usual determination and optimism. Having known George closely for 25 years, I am devastated by this loss. He was great in many respects: great father, great friend, great statistician, great collaborator, great researcher, great teacher, great editor, great runner. But, above all, he was a great and unique person. The loss is profound; the loss is significant, for me and for us. My thoughts go out to his wife, Anne, and children, Benjamin and Sarah, who are the ones who feel this loss the most keenly.

To me, George was the epitome of the academic researcher and a role model. He had great ideas, he was unbelievably enthusiastic about ongoing research, he was incredibly hard-working, he was an excellent co-author and co-editor, he had a good vision of what was going on in the field of statistics and of what would happen, he was always ready to embark on new directions of research, he was supportive of young researchers and students, he had superb organizational skills.

To take a few examples and make this statement more factual, consider the great job he did as a *Journal of the American Statistical Association* editor. As a reader, I think he improved the general quality of the journal (which was already high). As an associate editor, I can state he had a clear idea of the editorial line he wanted to follow and was helpful to authors in changing good papers into great papers. As an author, I can at last certify he was impartial (but fair) in dealing with my papers! His editorship of the *Journal of the Royal Statistical Society Series B*, which we shared for more than two years, was equally superb.

Consider also his successes with his PhD students, starting with Costas Goutis (who most sadly died in 1996): “George has always been my model as an adviser in that he simultaneously helped the students quite a lot from research topics to research methods to organization to preparation for academic careers and let them as free and autonomous as possible.”

Reflect on his involvement in *Environmetrics and Genetics*, his breadth in research topics, his numerous collaborations, and the number of grants he received and you get a unique picture.

I must add, from an even more personal point of view, that George was the epitome of the ideal man.



George Casella 1951–2012

Besides being a successful academic, he was, in parallel, a devoted father, a volunteer firefighter, a serious marathon runner, and a community servant. And all this with constant good cheer, attention to others, and a willingness to help whenever he could. I envied him for his relentless energy to lead so many lives at the same time so efficiently!

In a tribute to George’s most lasting legacy—his books—and following a terrific suggestion from *CHANCE*’s executive editor, Sam Behseta, I asked friends to write reviews of some of George’s most influential books. Hence, the special nature of this Book Reviews column: The books are not new publications, as some were written decades ago, and the reviews are not written by impartial reviewers, but by friends, in a memorial style. Nonetheless, I think they constitute a great collection of views on George’s books and their influence on the field, and thus serve to induce those who have not yet read them to do so and those who have read them to reassess them and their use in the classroom.

Andrew Gelman on *Introducing Monte Carlo Methods with R*

Christian Robert and George Casella



Year: 2007
Hardcover: 283+xix pages
Publisher: Springer-Verlag,
Use R! Series
ISBN-13: 978-1441915757

I remember being told many years ago that political ideologies fall not along a line, but on a circle: If you go far enough to the extremes, left-wing communists and right-wing fascists end up looking pretty similar.

I was reminded of this idea when reading Christian Robert and George Casella's fun new book, *Introducing Monte Carlo Methods with R*. I do most of my work in statistical methodology and applied statistics, but sometimes I back up my methodology with theory, or I have to develop computational tools for my applications. I tend to think of this sort of ordering:

Probability theory → Theoretical statistics → Statistical methodology → Applications → Computation

Seeing this book, in which two mathematical theorists write about computation, makes me want to loop this line into a circle. I knew this already—my own single true published theorem is about computation, after all—but I tend to forget. In some way, I think computation—more generally, numerical analysis—has taken some of the space in academic statistics that was formerly occupied by theorem-proving. It's great that many of our more mathematically minded probabilists and statisticians can follow their theoretical physicist colleagues and work on computational methods. I suspect applied researchers such as myself will get much more use out of theory applied to computation than to traditionally more prestigious work on asymptotic inference, uniform convergence, mapping the rejection regions of hypothesis tests, M-estimation, three-armed bandits, and the like.

Don't get me wrong; I'm not saying computation is the only useful domain for statistical theory. There are many new models to be built and limits to be understood. Just, for example, consider the challenges of using sample data to estimate properties of a network. Lots of good stuff to do all around.

Anyway, back to the book by Robert and Casella. It's a fun book, partly because they resist the impulse to explain everything or to try to be comprehensive. As a result, reading it requires the continual solution of little puzzles (as befits a book that introduces its chapters with quotations from detective novels). I'm not sure if this was intended, but it makes it a much more participatory experience, and I think it would be an excellent book for a course on statistical computing for that reason.

For instance, there is an example of optimization of the likelihood and log-likelihood on pages 127–128. However, it is never explained why these yield different optima, nor is the code actually given for the graphs that are displayed. Let me emphasize here that I am not stating this as a criticism; rather, Robert and Casella are usefully leaving some steps out for the reader to chew on and fill in.

I noticed a bunch of other examples of this sort, where the narrative just flows by and, as a reader, you have to stop and grab it. Lots of fun.

The good news is that there's an R package (mcsm) that comes with the book and includes all the code, so the interested reader can always go in there to find what they need.

One other thing: The book is not beautiful. It has an ugly mix of fonts, and many of the graphs are flat-out blurry. Numbers are presented to seven significant figures. Maybe that's okay, though, in that these displays look closer to what a student would get with raw computer output. The goal of the book is not to demonstrate ideal statistical practice (or even ideal programming practice), but to guide the student to a basic level of competence and give a sense of the many intellectual challenges involved in statistical computing. And this book does that well. The student can do what's in the book and is then well situated to move forward.

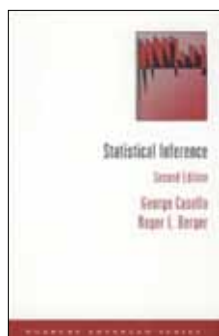
I think the book would benefit from a concluding chapter—or an epilogue or appendix—on good practice in statistical computation. Various choices are made for pedagogical reasons in earlier chapters that could, if uncorrected, leave a wrong impression in readers' minds. Beyond the aforementioned significant digits and ugly graphs, I'm thinking of choices such as the Langevin algorithm in Chapter 6 (which I understand has many practical problems and can most effectively be viewed as a special case of hybrid sampling) or the discussion of hierarchical models without the all-important (to me) redundant multiplicative parameterization. There's also the use of a unimodal distribution to approximate the likelihood function from Cauchy data and the overemphasis (from my perspective) of importance sampling, which is a great conceptual tool but is close to dominated by Metropolis-Hastings in practice. (As I wrote back in 1991, people view importance sampling as exact and MCMC as approximate, but importance sampling is not exact at all.) I recognize that the ideas of importance sampling, as applied to more complicated algorithms such as particle filtering and sequential

Monte Carlo, are important. I'm just less convinced of the relevance of straight importance sampling (of the sort discussed in the book), except as a way to introduce concepts that will become important later on.

In summary, there are many books about R that are intended as reference works. Robert and Casella's book is different: It's a short adventure that I think would be excellent to use as a textbook for students learning about statistical computing.

William E. Strawderman on *Statistical Inference*

George Casella and Roger L. Berger



Year: 2007

Hardcover: 283+xix pages

Publisher: Springer-Verlag,
Use R! Series

ISBN-13: 978-1441915757

I applaud the editor's decision to commemorate George Casella's contributions to the discipline of statistics and the lives and careers of his multitude of friends, colleagues, co-authors, and students through a series of reviews of his books. I am pleased to contribute to this memorial to my friend, colleague, and co-author.

I hired George for his first academic position as an assistant professor at Rutgers in 1977. He returned the favor by hiring my son, Rob, at Cornell in 2000. He has long been and will always be remembered as one of the people who have enriched my life with friendship, good fellowship, wonderful collegiality, and great humor. He was singularly energetic, ever optimistic, a wonderful teacher, and a caring mentor to students and young (and even not so young) colleagues. Many of us have been greatly blessed by his presence and deeply mourn his premature passing.

George's research output was broad and ranged from the highly theoretical to methodological developments to pure applications in various fields. Throughout his career, he was a valued consultant on numerous projects in a wide variety of subject areas. His substantial institutional and editorial contributions are also widely known and admired.

While many are aware of his statistical research, consulting activities, administrative and editorial activities, he is probably better known to the broad statistical community (and to students in particular) through his

numerous textbooks. He was a prolific and talented writer. His books are well written and have been well received by graduate students and practitioners. They cover a wide range, from relatively introductory to deep theory, from applied and theoretical linear models to experimental design and statistical computations. Most were written with co-authors and give evidence of his wonderful ability to interact and work with others.

This review is of what is probably the best known and most broadly used of his books. Casella and Berger is probably the most popular introductory statistical theory text for senior undergraduates and beginning graduate students in the United States and Canada, deservedly so, in my opinion. Somewhat ironically, while I often teach our upper-level PhD theory course out of Lehmann and Casella (pretty much my favorite course and my favorite book), I taught our first-year, two-semester PhD theory course out of Casella and Berger this year.

The book opens with a five-chapter introduction to probability theory and sampling distributions and then moves into the heart of the material on statistical theory, beginning with an introduction to likelihood, sufficiency, completeness, ancillarity, and equivariance in a chapter titled "Principles of Data Reduction." It moves on to chapters on point estimation (7), hypothesis testing (8), interval estimation (9), asymptotics (10), analysis of variance and simple linear regression (11), and finally, regression (12) (errors in variables, logistic and robust). There is an appendix on computer algebra, and a set of tables of common distributions.

The authors suggest a reasonable one-year course can cover most of chapters 1–10 with certain sections deleted, and this is consistent with my experience.

The chapters on probability give a nice introduction to basic concepts of probability theory, placing particular emphasis on those aspects that are key to statistical theory: discrete and continuous families of distributions, exponential families, moments and moment generating (but not characteristic) functions, univariate and multivariate change of variables, sampling distributions, the law of large numbers, and the central limit theorem.

These chapters do have a minor weakness shared by most books at this level. Unable to give a full measure theoretic treatment, the authors attempt to take advantage of at least some of the unifying aspects that measure theory brings to the development, but certain concepts are incompletely covered. The treatment of expected value exemplifies this, and the authors pay a small, and typically for George, humorous (even when he is not amused), tribute to the difficulty in the definition of expected value on Page 55. On the other hand, the treatment of interchange of integration and differentiation in Chapter 2 is an example of the advantages of this approach. The sections on inequalities and identities in chapters 3 and 4 are particularly nice.

Chapter 5, on sampling distribution, is a particularly nice bridge between probability and statistics and gives a modern twist to the discussion by introducing computational issues involved in generating samples from specific distributions, including accept-reject methods and basic MCMC methods. There is also a nice discussion of order statistics.

Chapter 6, “Principles of Data Reduction,” gives a nice introduction to sufficiency, ancillarity, and completeness; discusses my favorite theorem (Basus: OK, OK, Steins Lemma is terrific tool!); and provides an even-handed discussion of the likelihood principle and introductory discussion of equivariance.

The next three chapters (7, 8, and 9) on point estimation, hypothesis testing, and interval estimation, respectively, follow a somewhat common pattern. An initial section introduces the general topic and is followed by a section on methods (ad hoc, likelihood, and Bayes) for carrying out the statistical objective. The final section then discusses methods of evaluating the various possible procedures, including some discussion of loss functions and frequentist and Bayesian optimality properties. There are nice presentations of standard topics such as the Cramer-Rao Inequality, the Rao-Blackwell, Lehmann-Scheffe theorems, and the Neyman-Pearson Lemma. Additionally, there are nice modern touches, such as a discussion of the EM algorithm. There is also a nice discussion of union-intersection and intersection-union tests. The inclusion (intersection?) of this topic is probably largely due to George’s co-author and best friend from graduate student days at Purdue, Roger Berger, another of the world’s good guys and a true union-intersection expert.

The chapter on asymptotics (10) has sections on estimation and testing and interval estimation, as well as a nice introductory section on robustness, including a discussion of the asymptotic distribution of the median and Huber’s estimator. Once again, the modernity of the text is evidenced via the (re)introduction of the bootstrap as a method for calculating standard errors.

Chapter 11 on analysis of variance and simple linear regression does a nice job of introducing these critical basic models and developing the standard least squares-based estimators, tests, and confidence intervals, including F-tests, simultaneous confidence intervals, and BLUEs.

The final chapter (Regression Models) discusses errors in variables regression, logistic regression, and robust regression. The discussion of errors in variables regression is particularly nice and also somewhat rare at this level.

Two notable features of the text are the problem sets at the end of each chapter and the concluding sections of each chapter, titled “Miscellanea.” The problem sets are extensive, with a minimum of 31 in Chapter 12 and a maximum of 69 in Chapter 5 (and a robust median of 52.5). Aside from being silly, it would, of course, be mean (52) and at variance (140.1818) with a number

(0) of editorial policies of *CHANCE* to give the standard deviation (11.839840).

The authors have taken care to include a wide range of difficulties for each problem set and include a breadth of problem areas. The Miscellanea sections enrich and broaden the discussion of each chapter. They give evidence of the care the authors exhibited in the choice of topics in the individual chapters and the depth and breadth of the authors’ knowledge of the subject. They also give further evidence of their appreciation of and love for the art and craft of teaching statistical theory.

Jean-Louis Foulley on *Variance Components*

Shayle R. Searle, George Casella, and
Charles E. McCulloch



Year: 1992

Hardcover: xxiii+501 pages

Publisher: Wiley-Interscience

ISBN-13: 978-0471621621

This book is devoted to variance components. Although there have been many books about mixed model methodology since it was published 20 years ago, it remains an essential reading in the statistical literature as the most complete textbook on this topic in the linear case. It covers in great detail the two most important families of estimators of variance components, namely the quadratic estimators (ANOVA, Henderson’s methods, MINQUE, and dispersion-mean model), but also the maximum likelihood-based estimators either in their standard form (ML) or as residual maximum likelihood (REML).

It also provides all the relevant basic techniques pertaining to mixed model methodology, including best linear unbiased prediction (BLUP), Henderson’s mixed model equations (MME), and the expectation-maximization (EM) algorithm. It also gives some insight into estimation of variance components in the nonlinear case through the example of binary data.

There is no secret to the success and excellence of this book, since its three authors were eminent statisticians from the biometric unit of Cornell University, where most of these techniques were developed under

the guidance of Charles R. Henderson and his students and disciples.

The book consists of 12 chapters plus three appendices, one on special formulae for nested and two-way crossed classifications and the two others on results in matrix algebra and elementary statistics.

In Chapter 1 (Introduction), Searle, Casella, and McCulloch (SCM) defined basic terminology used such as factors, levels, cells, and effects; balanced and unbalanced data; and fixed and random effects with a variety of simple examples illustrating how to decide whether a set of effects is fixed or random. Since Eisenhart (1947), this remains one of the most difficult questions in specifying such models. SCM provide the reader with basic questioning on this issue such as are the levels of the factor randomly sampled from the distribution or are the effects attributable to finite sets that arise in the data because we are interested in them.

However, this distinction often remains ambiguous. For instance, are we talking about sampling levels or effects of a factor? The example of years is typical of this, with years levels obviously not random but year effects on yield of crops being likely to be unpredictable, at least in the short term. On the other hand, one may have a particular interest in some effects and still consider them as random.

This is the case of the favorite example of animal breeders with sire and herd-year used to analyze progeny data in the field, with sire treated by Henderson as random and herd-year as fixed, whereas the opposite would have been as much plausible. One way to circumvent this dilemma consists of referring to a Bayesian approach. It is too bad we have to wait until Chapter 9 on hierarchical models to hear that fixed and random effects are treated similarly and no distinction is made between them except possibly via their prior distributions.

Chapter 2 (History and Comment) is especially welcome, as it sets the historical context of variance components estimation. It recalls the main steps in the development of such methods starting from the pioneering works of Airy and Chauvenet in the second half of the 19th century up to the maximum likelihood procedures formalized by Hartley and Rao a century later and routinely applied nowadays. This publication marks a break with the quadratic era initiated by the work of Fisher on ANOVA and the intra-class coefficient (1925) and culminating with unbalanced data in Henderson's I, II, and III methods (1953) and Lamotte and Rao's MINQUE (1971) or equivalent.

I really enjoyed reading these 20 pages of history. They are extremely well documented and show how science proceeds as any evolutionary process along punctuated equilibria with sudden and abrupt jumps followed by longer periods of maturation and gradual increments.

Chapter 3 focuses entirely on the simplest example of the one-way classification, but treating it as completely as possible regarding data structures (balanced or not) and estimation procedures (ANOVA, F-statistics, ML, REML, and Bayes). These procedures are

re-examined in more detail in subsequent chapters. To that respect, being able to apply all these techniques to this simple model makes the reader ready to grasp more complex situations and understand the essence of mixed model methodology.

This is also an inexhaustible source (or treasure) of exercises for teachers and students. I especially like the example of how to estimate (predict) the IQ of college freshman Ronnie Fisher from the average of n IQ test scores by making use of the conditional mean. It reminded me of how Charles Henderson (Henderson, 1973b) discovered BLUP after also facing a deceptively simple problem assigned to his mathematics statistics class. The problem was given an IQ score of 130, what is the ML estimate of an individual true IQ? The moral of this story, if any, might be that deceptively simple problems are sometimes worthwhile to justify all the time and effort spent on them.

What also strikes the reader is how the transition from balanced to unbalanced data alters the nice distributional properties of ANOVA estimators making derivations of exact confidence intervals on variance components intractable. I suspect some readers might be greatly disappointed by subchapter 3.9 on Bayes estimation of variance components for this simple one-way model, ending up with intractable analytical results for the joint posterior distribution of variance components as well as for their modal values, even in the balanced case.

Chapter 4 is completely devoted to balanced data as defined by the authors, wherein all elementary cells (highest combination of levels of the different factors) have equal numbers of observations. One may wonder why the authors spend 55 pages on such an issue. Is it not really too much? The authors justify it by emphasizing the importance and attractiveness of well-designed experiments, which make, in many cases, ANOVA estimators of variance components having optimum statistical properties.

Personally, I see some practical reasons for being aware of such results. Most routine procedures for linear mixed models are based only on techniques derived for unbalanced data so that the nice properties of the statistics (orthogonality of mean squares, appropriate F-statistics, exact p -values) remain mistakenly hidden in the outputs when applied to balanced designs.

Chapter 5 deals with ANOVA-type estimators of variance components in the unbalanced case. After reminding us of the basic principles of such moment quadratic estimators, the chapter is almost entirely devoted to Henderson's works in this area. It is only fair that a detailed account of Henderson's methods I, II, and III are presented here. Henderson's 1953 paper in *Biometrics* marked a breakthrough in the area of variance component estimation for unbalanced data. He was a genius to capitalize on the knowledge of ANOVA techniques in the balanced case to transfer them appropriately to the unbalanced case. His methods were easy to understand and compute (at least

the two first), even to large data sets. Estimators are obtained by equating a set of quadratic forms to their theoretical expectations under the model (random or mixed) considered for the analysis.

Henderson's method III is the most accomplished among the three proposed. It uses quadratic forms derived from fitting by least squares different submodels. But computations of the expectations of quadratic forms can require the inversion of large matrices, and this was a real drawback, limiting the application of this method to toy examples by the time it was proposed.

A key and topical question is discussed by SCM at the end of this chapter about comparing different methods. SCM emphasize that the ANOV- type methodology, itself, gives no guidance whatever as to which set of quadratic forms is, or might be, optimal in any sense. Unfortunately for the practical user, they conclude that a fair comparison is virtually unfeasible, as the sampling variances of estimators depend on too many combinations of parameters and data patterns.

Chapter 6 provides us, in a relatively shorter space, with the general theory of maximum likelihood estimations of parameters in linear mixed models in a comprehensive form. Several important aspects were outlined by SCM, such as the constraints of maximizing the likelihood within the parameter space and kindred numerical issues (iterative scheme and convergence problems to local or global maxima). The two-way crossed random model with and without interaction is treated analytically in full detail.

The end of the chapter tackles restricted (or better residual) maximum likelihood (REML) so as to correct for bias arising in ML by not taking into account the degrees of freedom used for estimating fixed effects. The coverage is neat, but to my taste too short. REML is introduced via the likelihood of the so-called error contrasts according to Harville's terminology (i.e., residuals obtained after fitting fixed effects by ordinary LS). I would have liked to see alternative angles of attack such as i) by conditioning and factorizing the likelihood into two parts with one depending only on dispersion parameters (see Kalbfleisch and Sprott, 1970) and ii) by maximizing a marginal likelihood obtained by integrating out fixed effects with respect to a flat prior (Harville, 1974). In fact, this last interpretation appears only 75 pages farther in Chapter 9 on hierarchical models. Anyway, this last method is especially important to understand an EM version of REM treating fixed effects as part of the missing data vector (see Dempster, Laird, and Rubin, 1977).

Another important issue that has not been covered lies in hypothesis testing about variance components for values located on the boundary of the parameter space and which raises some nasty complications.

What a delight to read Chapter 7 on prediction of random variables! This is probably the most comprehensive account of this subject in the statistical literature. It starts with the exercise on prediction of individual IQ based on observed scores by Mood that

inspired, as seen previously, Henderson at the beginning of his career. It reviews the different methods of prediction, namely best prediction (BP), best linear prediction (BLP), and best linear unbiased prediction (BLUP). A clear distinction is introduced between estimating parameters and predicting random variables, especially as far as properties such as expectation are concerned.

A substantial subchapter is inserted on Henderson's mixed model equations (MME), which simultaneously yield GLS estimations of fixed effects and BLUP of random effects and numerous byproducts, including ingredients for EM and EM-like algorithms for ML and REML estimations of variance components. This was a major contribution by Henderson to mixed model methodology, both in terms of computing efficiency and brilliant interpretations of solutions (see hierarchical Bayes models and shrinkage estimations), which has remained unknown for too long by the academic statistical community. With the advent of rating and ranking as a prominent domain of application of statistics, it was a fair initiative of SCM to remember BLUP and MME as a key reference in their textbook.

Chapter 8 is about numerical methods and issues in computing ML and REML. Maximizing complex nonlinear functions of parameters is, per se, a difficult problem due to the existence of stationary points and extrema. When, in addition, this optimization involves constraints on the parameter space, it becomes even harder. For instance, how to cope with maxima occurring on the boundary of the parameter space? SCM review two kinds of omnibus iterative techniques of optimization: i) methods based on first and second derivatives such as Fisher scoring, Newton-Raphson and Marquardt's methods, and ii) EM procedures.

The first ones have clearly shown their efficiency, as they have been taken up by most computing packages. I was a little bit disappointed by the way SCM presented the EM algorithm for computing ML and REML estimates of variance components. First, they do not take advantage of the properties of Henderson's MME, which are so convenient in that case and lead to formulae easier to handle than those based on the inverse of V (the variance covariance matrix of the data vector). Moreover, these EM formulae have close similarities with expressions (68ab and 91ab) given in subsections 7.6.cd, but also some differences that deserve more attention.

Their second EM algorithm (8.3.d), made by using a GLS estimate of the fixed effects at the end of the iteration process is, as they mentioned it, not an EM algorithm, but what was called later an ECME (expectation conditional maximization either) algorithm by Liu and Rubin (1994). I also regret that they do not elaborate on the EM algorithm that takes fixed effects as part of the missing data vector in addition to the random effects. They just drop a hint about it in Chapter 9, Section 2b, thus REML estimation is estimation that has the values of both β and u integrated out, but out of the EM context.

Chapter 9 ends up with an analytical illustration of the one-way random model (very useful as a source of exercises) and a too brief overview of standard computing packages (Genstat, SAS, BMDP). A special mention should be given nowadays to AS-REML by Gilmour, Thompson, and Cullis (1995), which is based on the hybrid solution of averaging the expected and observed information matrices in the iterative system of REML equations.

Chapter 9 explores another approach to the analysis of mixed models consisting of hierarchical modeling and Bayesian inference. It is first shown how the general mixed model expression can be formulated as a two- or three-level hierarchy. It also gives an interpretation of ordinary and residual likelihoods according to the assumptions made on the prior distributions of fixed effects (point, mass, and uniform distributions, respectively). In the normal conjugate case, it establishes links between Bayesian point estimators based on posterior distributions of fixed (β) and random (u) effects and their classical counterparts, namely GLS and BLUP, respectively, assuming a prior on u centered at zero and prior on β with infinite variance.

Empirical Bayes estimation is also outlined, emphasizing the danger of the substitution principle for estimating the sampling variance of estimators. One way to overcome these difficulties is to adopt the strategy of Kass-Steffey that SCM advocate to obtain reasonable variance approximations. Other types of hierarchies outside the normal framework are presented (e.g., the beta-binomial and logit-normal cases). A technique for calculating ML or MAP estimations of parameters is presented in subchapter 9.5 (pages 350–351) based on what SCM call hierarchical EM. I am not sure I really captured the essence of this short-cut procedure. For instance, in the example of linear normal mixed models, in the M step, we not only need the conditional mean of random effects (u) given the data and parameters at their current values, but also their conditional variance.

Finally, the authors outline the great merits of hierarchical modeling, along with kindred Bayesian procedures for the users both conceptually and technically. We have not to worry about what quantities are fixed or random (\cdot). We have only to worry about whether the quantity is observable (data) or unobservable (parameter) and calculating the distribution of the unobservable given (conditional on) the observable. Is there a better conclusion to summarize the philosophy of this chapter?

Chapter 10 is about binary and discrete data. This is a short chapter (10 pages) as compared to the previous ones. Actually, it comes back to the same models as those presented in Section 4 (other types of hierarchies) of Chapter 9 and remains restricted to binary data despite its title. SCM review the three standard models prevailing for such data (i.e., the beta-binomial, logit-normal, and probit-normal models). Their merits and drawbacks are well discussed, especially the limitations of the first one, which precludes any kind of regression modeling with covariates specific to elementary responses.

Chapter 11 is titled “Other Procedures.” It is too bad that this chapter looks at a real hotchpotch of different topics and techniques such as i) modeling variance and covariance components in multidimensional data structures, and ii) alternative methods of estimating variance components.

The first part on modeling covariance is highly welcome, although more explicit and practical examples would help the users clarify the modeling issues. I especially think about variance covariance structures of models involving different traits and time measurements such as the unstructured AR(1) pattern.

The second part, modeling variance components as covariances, is not completely useless, but does not deserve such a long development (maybe an exercise).

The third part, criterion-based procedures, is well treated, but definitively ill positioned in the book. Lamotte and Rao’s MINQUE procedures should have been located somewhere between Henderson’s methods (Chapter 5) and ML and REML (Chapter 6), as these make a clear link between these two approaches.

Chapter 12 deals with an approach that is rather unusual in textbooks about mixed models, the so-called dispersion-mean model due to Pukelsheim. What is it? In this model, the data vector is made of some translation invariant forms (squares and cross-products of OLS response residuals), which can be expressed as a linear model of the vector of variance components. It is then shown that ordinary least squares equations applied to this model are the MINQUE0 equations and that GLS yields REML equations under normality.

Interestingly, as pointed out by SCM, the same approach can be used to extend REML for non-normal data, and this might be an alternative to marginalized likelihood (in the Bayesian sense). To my knowledge, this has not yet been applied.

In conclusion, it turns out that *Variance Components* is not only a major textbook on a topical subject, but also a mandatory one for all statisticians willing to learn the basics of linear mixed models. Having been published 20 years, it might benefit from a new edition with updated material, especially on generalized and nonlinear mixed models and kindred Monte Carlo techniques, both under the frequentist and Bayesian frameworks. That also would provide an opportunity to reorganize the contents of the book. In any case, it is, and will be, a classic for a very long time, and you’d better have it on your shelves if you want to use and/or say something about mixed models and variance components.

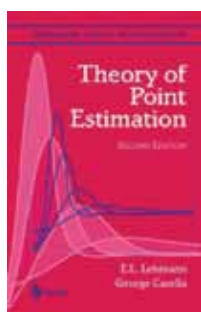
Further Reading

- Arville, D. A. 1974. Bayesian inference for variance components using only error contrasts. *Biometrika* 61:383–385.
- Gilmour, A., R. Thompson, and B. Cullis. 1995. Average information REML, an efficient algorithm for variance parameter information in linear mixed models. *Biometrics* 51:1440–1450.

- Kalbfleisch, J. D., and D. A. Sprott. 1970. Application of the likelihood methods to models involving large numbers of parameters. *J Royal Statistical Society B* 32:175–208.
- Liu, C., and D. B. Rubin. 1994. The ECME algorithm: A simple extension of the EM and ECM with faster monotone convergence. *Biometrika* 81:633–648.

Larry Wasserman on *Theory of Point Estimation: Second Edition*

Erich Lehman and George Casella



Year: 1998
Hardcover: 616 pages
Publisher: Springer-Verlag
ISBN-13: 978-0387985022

What happens when one of the most gifted writers in the field of statistics asks another gifted writer to help him write a second edition of his book? And not just any book. The book happens to be a classic. The result is *Theory of Point Estimation* (2nd edition) by Erich Lehmann and George Casella.

Erich Lehmann wrote *Theory of Point Estimation* in 1983. It quickly became a standard for PhD courses in theoretical statistics. For generations of statisticians, *Theory of Point Estimation* defined the core of statistical theory. Passing a qualifying exam meant mastering the contents of the book.

Why did Lehmann's book become so important? Perhaps there was a need for a book at just this level. Perhaps it was the selection of topics, which made it just right for so many PhD programs. But I think the most important factor is the writing. Lehmann had a knack for covering difficult topics with unusual clarity and economy. A good example is the section on measure theory, which manages to condense the essential topics into a mere 12 pages.

The second edition came out 15 years later, in 1998. Why did Lehmann ask George Casella to help him write it? The answer is obvious: George had established himself as another Lehmannesque writer, another

statistician with the gift of writing exceptionally clear expositions. Indeed, *Statistical Inference* by George Casella and Roger Berger is another classic, widely used throughout the world.

I remember George telling me that when Lehmann asked him to collaborate on the second edition he was flattered but also a bit intimidated. How do you update a classic? The approach they chose was wise. A drastic re-writing was out of the question. Instead, they decided to preserve most of the book and update the text by adding new material that reflected much of what had happened in statistics between 1983 and 1998. In particular, the added material reflected George's increasing attention to Bayesian inference and posterior simulation.

The original edition consists of six chapters: Preparations, Unbiasedness, Equivariance, Global Properties, Large Sample Theory, and Asymptotic Optimality. Even without updating, the first edition holds up well today. The material in the chapters on unbiasedness and equivariance has become less relevant, but the remainder is still crucial. Every statistician needs to be familiar with minimax theory, shrinkage, Bayes estimators, convergence, and asymptotic efficiency. There are, of course, many other treatments of these topics today. But anyone wanting a clear understanding of the essentials would do well to read these chapters.

So what changed in the second edition? The most significant changes are found in chapters 4 and 5. Chapter 4 is now called Average Risk Optimality and brings modern Bayesian inference into the picture. In particular, the chapter contains sections on hierarchical Bayes and empirical Bayes. The discussion of hierarchical Bayes contains a succinct introduction of Gibbs sampling, which is a must for any modern treatment of the subject. It also contains some information-theoretic ideas. For example, there is a proof, using Kullback-Leibler distances, that the posterior of a hyper-parameter is less sensitive to choice of prior than the posterior of a parameter. (This relates to work by Goel and DeGroot in the early 80s that should be better known.) There is also discussion of reference priors and a statement of an elegant theorem by Clarke and Barron (1990) about the Kullback-Leibler distance between the prior and posterior.

The subsection on empirical Bayes is replete with examples and even has an introduction to robust Bayesian inference.

The original fourth chapter (Global Properties) on minimax theory and admissibility is now Chapter 5 (Minimaxity and Admissibility). The material on shrinkage estimation has been expanded and includes, for example, the role of superharmonic functions and minimaxity.

Chapter six (Asymptotic Optimality) now begins with an introductory subsection, giving the reader some preparation before jumping into the main details.

The only major deletion that I am aware of is the removal of the material on robust estimation. This makes good sense. Topics like L and R estimators do not command the same attention today as they did in 1983.

There are numerous small changes as well. For example, there are more exercises and the references are expanded and put at the end of the book. There is a section called Notes at the end of the chapter with extra topics and historical perspective. There are lots of interesting nuggets here such as curved exponential families, large deviation theory, weak differentiability, the ergodic theorem, the Hunt-Stein theorem, and estimating equations, to name a few.

Reviewing Lehmann and Casella is a bittersweet experience. Looking back at the book, it was wonderful to see two masterful writers at work. The book is a testament to the power of clear writing. And seeing a classic updated and improved after a 15-year gap is fascinating. But it is a sad reminder that we lost two great statisticians, Lehmann in 2009 and Casella in 2012.

Ironically, we are approaching the 15-year mark since the publication of the second edition. Who could possibly do yet another update? Can anyone fill the shoes of these singular expositors? I think not. Perhaps we will have to content ourselves with the fact that the second edition may be the last. Keep it on your shelf and cherish it.

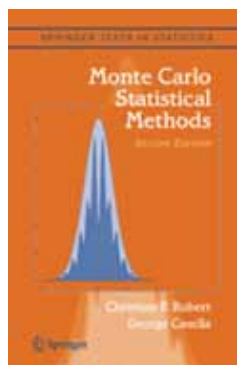
Further Reading

Clarke, B. S., and A. R. Barron. 1990. Information-theoretic asymptotics of Bayes methods. *IEEE Transactions on Information Theory* 36: 453–471.

Goel, P. K., and M. H. Degroot. 1981. Information about hyperparameters in hierarchical models. *Journal of the American Statistical Association* 140–147.

Xiao-Li Meng on Monte Carlo Statistical Methods

Christian Robert and George Casella



Year: 2004
Hardcover: xxx+645 pages
 (2nd. edition) 18
Publisher: Springer-Verlag
ISBN-13: 978-1441919397

Xiao-Li, what are you talking about? We do not ask people who are not overcommitted!

~ George Casella

All textbooks George Casella has co-authored that I have seen started each chapter with a quote. Given that these textbooks were written with different coauthors, a reasonable inference is that George was the one who instituted the tradition, and perhaps also selected most of the quotes. Although the contextual link between George's selection and the corresponding chapter sometimes requires deep reflection, the quote above largely motivated this book review.

Almost a decade ago, George called and asked me to serve as the editor of a major journal. Very honored, I nevertheless declined, citing over-commitment, having just been appointed department chair. The quote was George's loud and emphatic response. To George, of course, there was no such thing as over-commitment. He was a living example of the "The Hilbert Hotel"—there is always one more room for a newly arrived commitment.

History tends to repeat. When Christian Robert asked me to write a book review for this special collection in memory of George, I was just given a deanship. Intriguingly, the difference between writing a book review and editing a major journal is not an entirely inappropriate analogy for comparing the roles of a department chair and a dean with 57 (and still counting) departments and programs to worry about. But how could I possibly say no, with George's emphatic response still ringing in my ear?

However, I have never written a book review or a book. I did try both, but in the end a book was always too heavy, regardless of being an author or a reviewer. How could I possibly then pick up this one, particularly after weeks of receiving a heavy dose of dean's meetings? What I needed badly was a weekend retreat, not a weekend review! With the help of a glass of Two Hands and the rhythm of the ever-intoxicating Ebru Gündeş, I nevertheless sat down and opened my complimentary copy of *Monte Carlo Statistical Methods*. I was quite aware of its good reputation as a graduate-level textbook, but my mood then demanded a bit more. An intellectual massage perhaps would not relax me as much as a physical one, but surely it would help to taper my desire to internalize all these mysterious Turkish lyrics!

The first five chapters turned out to be a rather soothing introduction to the world of Monte Carlo: (1) Introduction; (2) Random Variable Generation; (3) Monte Carlo Integration; (4) Controlling Monte Carlo Variance; and (5) Monte Carlo Optimization. The writing is both concise and informative, with worked-out examples following immediately after most concepts, theory, or methods. There are ample exercises for each chapter, followed by Notes, which really are intellectual desserts, treating those still hungry for more food for thought, even after going through the regular material and many homework problems.

I was particularly pleased to see the chapter on Monte Carlo Optimization, not because it includes Monte Carlo EM (which, actually, is not a delicacy for

me, even though I have helped to create a few EM-type recipes). Rather, the vast majority of Monte Carlo treatments in statistics have been occupied by sampling and integration, to a point that many students are not aware of any other purpose for getting Monte Carlo samples. It was therefore refreshing to see Monte Carlo Optimization on the menu!

The next five chapters provide a rather paved—but by no means short—path into the kingdom of MCMC, that is, Markov Chain (or More Complicated!) Monte Carlo. All the essential theoretical navigation maps and guides, at least for first-time tourists, are given in the 60 pages of Chapter 6, Markov Chains, which could be viewed as a mini CliffsNotes of the authoritative account on this subject: Sean Meyn and Richard Tweedie's *Markov Chains and Stochastic Stability*.

Chapters 7, 9, and 10, respectively, detail the popular Metropolis-Hastings Algorithm, The Two-Stage Gibbs Sampler, and The Multi-Stage Gibbs Sampler. Although I was slightly misled initially by the titles of chapters 9 and 10 because of their usage of “stage” instead of the more common (and appropriate) “step,” I have no trouble recommending them to anyone who wishes to familiarize themselves with the recipes of these popular algorithms as well as their culinary principles and origins.

Whereas many other authors (me included) are likely to incorporate Chapter 8, The Slice Sampler, into the Gibbs sampler chapters as a special case, I can see the pedagogical rationale for introducing it before presenting the general Gibbs sampler. Throughout the textbook, the evidence is overwhelming that the authors care deeply about making the learning path as gradual and paved as possible.

About the Author

Christian P. Robert is a professor of statistics at Université Paris-Dauphine, France. He has written extensively about Bayesian statistics and computational methods, including the books *The Bayesian Choice* and *Monte Carlo Statistical Methods*. He has served as president of the International Society for Bayesian Analysis and editor in chief of the *Journal of the Royal Statistical Society (Series B)*.

With perhaps the exception of Chapter 12, Diagnosing Convergence, the rest of the book is devoted to materials that are less palatable for first-timers. Indeed, when I was deciding which chapters I should read most carefully (given I can afford only one weekend retreat), Chapter 11, Variable Dimension Models and Reversible Jump Algorithms, and Chapter 13, Perfect Sampling, were my first choices. This is because, although I have visited the MCMC kingdom many times, these two sites still induce an adventurous feeling every time I pass by. Not surprisingly, the author's skillful presentations helped to reduce my anxieties, perhaps permanently. I didn't have time to enjoy the last chapter, Chapter 14, Iterated and Sequential Importance Sampling, but if I did, I have little doubt I would have experienced the same feeling.

There were, of course, minor imperfections here and there, like any book ever or to be written. Overall, this is a highly recommended textbook for an introductory-to-intermediate level course on Monte Carlo, as well as an easily accessible reference book. It strikes a skillful balance between being concise and being comprehensive, with enough menu items to choose from without any being too heavy or unhealthy.

If there is anything that can be improved upon, it is simply something that faces every book—one cannot include the updates and advances developed after the publication of the book. But most of the new developments (e.g., on perfect sampling) do not change much of the materials in the first 10 chapters; for recently developed recipes, one can consult the *Handbook of Markov Chain Monte Carlo* (Brooks, Gelman, Jones, and Meng, 2011). Therefore, even with possible updates in mind, I would still recommend this edition for most people's bookshelves. Why not everyone? Well, I must sell a few copies of our handbook as well, so I can afford a real Monte Carlo retreat, even though this simulated one was far more relaxing than I initially expected, at least intellectually!

Editor's Note: Contributions in George Casella's name can be made to a fund at Purdue University. Send all correspondence to Rebecca Doerge, Department of Statistics, Purdue University, West Lafayette, IN 47907.

Become a part of the statistical community by joining
the American Statistical Association today.

Find the membership that is right for you at
www.amstat.org/join!