Book reviews, Spring 2013 . Special issue on George Casella's books

Jean-Louis Foulley, Andrew Gelman, Xiao-Li Meng, William Strawderman, Larry Wasserman, Christian Robert

Book reviews, Spring 2013
Special issue on George Casella's books

Jean-Louis Foulley
I3M - Institut de Mathmatiques et de Modlisation de Montpellier,
Andrew Gelman
Department of Statistics, Columbia University
Xiao-Li Meng
Department of Statistics, Harvard University
William Strawderman
Department of Statistics, Rutgers University
Larry Wasserman
Department of Statistics, Carnegie Mellon University
Christian P. Robert
Université Paris-Dauphine, CEREMADE, IUF, and CREST

December 10, 2012

Abstract

This note is made of an obituary of George Casella by Christian Robert and of five reviews of books written by George Casella, books that appeared a while ago, as a tribute to our friend and colleague and to his influential impact on the field. Those reviews are written by Andrew Gelman, William Strawderman, Jean-Louis Foulley, Larry Wasserman, and Xiao-Li Meng, respectively. The note is scheduled to appear in the next issue of *CHANCE*.

George Casella, 1951–2012: an obituary

On June 17 2012, my dear friend George Casella passed away, after a long illness he had fought with his usual determination and optimism. Having known George quite closely for 25 years, I was and am devastated by this loss. He was a great in many respects: great father, friend, statistician, collaborator, researcher, teacher, editor, runner, but, above all, a great and unique person. The loss is profound, the loss is significant, for me and for the community. My thoughts go out to his wife, Anne, and children, Benjamin and Sarah, who are the ones to 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 on-going research, he was incredibly hard-working, he was an excellent co-author and co-editor, he had a very 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 very supportive of young researchers and students, he had superb organisational skills, and so on.

To take a few examples (beyond the books!) and make this statement more factual, consider the great job he did as an editor of *JASA*: as a reader, I think that he improved the general quality of the journal (which was already high); as an Associate Editor, I can also state that he had a very clear idea of the editorial line he wanted to follow and was very helpful to authors in changing good papers into great papers; as an author, I can at last certify that he was quite impartial (but fair) in dealing with my papers! [His editorship of *JRSS Series B*, which we shared for more than two years, was equally superb!] Consider also the success stories with his Ph.D. students, starting with our friend Costas Goutis (who most sadly died in 1996 in a diving accident): George has always been my model as an advisor 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. Consider yet again his involvement in environmetrics and genetics, his breadth in research topics, his numerous collaborations, the number of grants he got, and you get quite a unique picture!

I must add, from an even more personal point of view, that George was also the epitome of the ideal man: besides being a successful academic, he was in parallel a devoted father, a volunteer firefighter, a serious marathon runner (who actually got me into running when we met in Purdue in the late 1980's!), while being also involved in community service. And all this with a constant good cheer, a permanent attention to others, an on-going willingness to help whenever and whoever he could. I always 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 academic legacy, namely his books, and following a terrific suggestion from Sam Behseta, to whom I am deeply grateful, I have asked friends to write reviews on the most influential books written by George. Hence the unusual features of this *Book Reviews* tribune: the books are not new publications as some have been written decades ago, the reviews are not written by impartial reviewers but by friends, in a memorial style rather than trying to assess potential usages and flaws of the books. Nonetheless, I think they constitute a great collection of views and perspectives on George's book, on their influence on the field, and thus serve a role beyond the tribute,[1] namely to induce those who had not yet read some of those books by George to read them and those who had already read them to re-assess them and use them in the classroom. Hence acting as genuine book reviews *per se*!

Andrew Gelman on "Introducing Monte Carlo Methods with R" by Christian Robert and George Casella

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

I remember many years ago being told 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 all about computation, makes me want to loop this line in 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 that computation–more generally, numerical analysis–has taken some of the place in academic statistics that was formerly occupied by theorem-proving. It's great that many of our more mathematical-minded probabilists and statisticians can follow their theoretical physicist colleagues and work on computational methods.

[1] A Memorial Fund has been set by Purdue University, where George graduated, in his memory.

I suspect that applied researchers such as myself will get much more use out of theory as applied to computation, as compared 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 that computation is the only useful domain for statistical theory, or anything close to that. There are lots of new models to be built and lots of 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 the book 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 for that reason it would also be an excellent book for a course on statistical computing.

For an example of an ambiguity or puzzle: on pages 127-128, there is an example of optimization of the likelihood and log-likelihood. 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 over 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, and the authors are also preparing a document with solutions to all their exercises.

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 7 significant figures. Maybe that's ok, 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 rather to guide the student to a basic level of competence and to give a sense of the many intellectual challenges involved in statistical computing. And that, this book does well. The student can do what's in the book and then is well situated to move forward from there.

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 lots of 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; or the use of a unimodal distribution to approximate the likelihood function from Cauchy data; or 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, for some reason 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 a lot of books about R that, some way or another, 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 to learn and think about statistical computing.

William E. Strawderman on "Statistical Inference" by George Casella and Roger L. Berger

- **Year:** 2001
- **Hardcover:** 660 pages
- **Publisher:** Duxbury Advanced Series (2nd edition)
- **ISBN-13:** 978-0534243128

I applaud the editors' decision to commemorate George Casella's contributions to the discipline of Statistics and to 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 familial 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 good 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 very 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 very 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 US 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), this year, I taught our first year, two semester PhD theory course out of Casella and Berger.

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 entitled Principles of Data Reduction (6). 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 that a reasonable one year course can cover most of chapters 1 - 10 with certain sections deleted and this is consistent with my experience this past academic year.

The chapters on Probability give a nice introduction to basic concepts of Probability Theory, placing particular emphasis on those aspects which 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 number 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 some computational issues involved in generating samples from specific distributions, including accept-reject methods and basic MCMC methods. There is also a very nice discussion of order statistics.

Chapter 6, Principles of Data Reduction, gives a nice introduction to Sufficiency, Ancillarity and Completeness, discusses my favorite theorem (Basu's: OK, OK, Stein's Lemma is terrific too!), and gives an even handed discussion of the likelihood principle, and an 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 of frequentist and Bayesian optimality properties. There are very nice presentations of the standard topics, such as the Cramer-Rao Inequality, the Rao-Blackwell, and Lehmann-Scheffé theorems, and the Neyman-Pearson Lemma. Additionally, there are some 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, testing and interval estimation as well as a nice introductory section on robustness, including a discussion of the asymptotic distribution of the median and of 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 BLUE's.

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 is 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 entitled Miscellania. 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 to include a breadth of problem areas. The Miscellania sections enrich and broaden the discussion of each chapter. They give evidence of the care that the authors have 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" by 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 on mixed model methodology since it was published twenty 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 non linear case through the example of binary data.

There is no secret in 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 twelve 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 Mc Culloch (later on referred to as SCM) defined the basic terminology used such as factors, levels, cells and effects; balanced and unbalanced data; 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 question 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 a finite sets that arise in the data because we are interested in them". However, this distinction remains often 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 on the short term. On the other hand, one may have a particular interest in some effects and still considering 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 that 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 us the main steps in the development of such methods starting from the pioneering works of Airy and Chauvenet in the second half of the nineteenth century up to the maximum likelihood procedures that were formalized by Hartley and Rao, one century later and which are routinely applied nowadays. This publication marks a break with the quadratic era initiated with the work of Fisher on ANOVA and the intra class coefficient (1925) and culminating for unbalanced data in Henderson's I, II and III methods (1953) and in Lamotte's and Rao's MINQUE (1971) or equivalent.

I really enjoyed reading these twenty pages of history that are extremely well documented and which 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 1-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 will be re-examined in more details 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 on 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. This reminded me of how Charles Henderson (Henderson, 1973b) discovered BLUP after also facing "a deceptively simple problem assigned by Mood 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 eg 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 1 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 the case of 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 ask oneself why spending 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 to be 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 remembering us 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 by the authors as yet no other books did it. Henderson's 1953 paper in Biometrics marked a breakthrough in the area of variance component estimation for unbalanced data. Henderson has had 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 to 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 method among the three proposed. It uses quadratic forms derived from fitting by least squares different sub-models. But computations of the expectations of quadratic forms can require the inversion of large size 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 on how comparing different methods. SCM emphasize that the ANOVA type methodology itself gives no guidance whatever as to which set of quadratic forms is, or might be, optima in any sense. Unfortunately for the practical user, they conclude that a fair comparison is "virtually not feasible" as the sampling variances of estimators depend of too many combinations of parameters and data patterns.

Chapter 6 provides us, in a relatively shorter space than for the previous ones, with the general theory of maximum likelihood estimations of parameters in linear mixed models in a very comprehensive form. Several important aspects have been 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 2-way crossed random model with and without interaction is treated analytically in full details.

The end of the chapter tackles restricted (or better residual) maximum likelihood (REML) so as "to correct for some bias arising in ML by not taking into account of 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 ie 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 of 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 further in chapter 9 on hierarchical models. Anyway, this last method is especially important in order to understand an EM version of REM treating fixed effects as part of the missing data vector (see Dempster, Laird and Rubin (1977) in their landmark paper on EM). 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 the next 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, CR 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 some properties are concerned such as expectation.

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 by-products including ingredients for EM and EM-like algorithms for ML and REML estimations of variance components. This was a major contribution of CR Henderson to mixed model methodology both in terms of computing efficiency but also in brilliant interpretations of solutions (see hierarchical Bayes models and shrinkage estimations) which has been unfortunately remained unknown for a too long time 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 us BLUP and MME as a key reference in their text book.

Chapter 8 is about numerical methods and issues in computing ML and REML. Maximizing complex non linear functions of parameters is per se a difficult problem due in particular to the existence of different kinds 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's 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 present 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 on using at the end of the iteration process a GLS estimate of the fixed effects is as they mentioned it, not an EM algorithm but what was called later on an ECME (Expectation Conditional Maximization Either) algorithm by Liu and Rubin (1994). I also regret that they do not elaborate more explicitly 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 beta and u integrated out", but out of the EM context.

Chapter 9 ends up with an analytical illustration to the 1-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 2 or 3 levels 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" (beta) and "random"(u) effects and their classical counterparts namely GLS and BLUP respectively assuming a prior on u centered at zero and prior on beta 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 eg the beta-binomial and the 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 the 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 () We have only to worry about whether the quantity is observable (data) or unobservable (parameter) and worry about 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 very 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 ie the beta-binomial, the 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 entitled "Other procedures". It is too bad that this chapter looks as a real hotchpotch of so 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 to clarify the modeling issues. I especially think about variance covariance structures of models involving different traits and time measurements such as, for instance, 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 (may be an exercise). The third part entitled "Criterion-based procedures" is very 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 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 it is also a mandatory one for all statisticians willing to learn the basics in linear mixed models. Having been published 20 years, it might benefit from a new edition with updated material especially on generalized and non linear mixed models and kindred Monte-Carlo techniques both under the frequentist and Bayesian frameworks. That might be also 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 better have it on your shelves if you want to use and/or say something about mixed models and variance components.

Further references

Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statist. Science*, **7** 457–511.

Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385

Gilmour, A., Thompson, R., and Cullis, B. (1995). Average information REML, an efficient algorithm for variance parameter information in linear mixed models. *Biometrics*, **51**, 1440–1450

Kalbfleisch J.D. and Sprott, D.A. (1970). Application of the likelihood methods to models involving large numbers of parameters. *J. Royal Statistical Society B* **32**, 175–208

Liu C. and Rubin D.B. (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" by 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: Second Edition* by Erich Lehmann and George Casella.

Erich Lehmann wrote *Theory of Point Estimation* in 1983. It quickly became a standard for Ph.D. 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 Ph.D. 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 the second edition? 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 to 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 of the material 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 Chapter 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 80's 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 the 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, 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 references

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

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

Xiao-Li Meng on "Monte Carlo Statistical Methods" by Christian Robert and George Casella

- **Year:** 2004
- **Hardcover:** xxx+645 pages (2nd. edition)
- **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 and I have seen started each chapter with a quote. Given that these textbooks were written with different co-authors, 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 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 of 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 need badly is a weekend retreat, not a weekend review! With the help of a glass of "two hands" and the rhythm of the ever intoxicating Ebru Gnde, I nevertheless sat down and opened my complimentary copy of "Monte Carlo

Statistical Methods" (Robert and Casella, 2004). 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 mysteries Turkish lyrics!

The first five chapters turn 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-page Chapter 6, Markov Chains, which could be viewed as a mini "CliffsNotes" of the authoritative account on this subject, Meyn and Tweedie (1993).

Chapters 7, 9 and 10, respectively, detail the popular Metropolis-Hastings Algorithm, The Two-Stage Gibbs Sample, and The Multi-Stage Gibbs Sampler. Although initially I was slightly misled 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 (myself included) are likely to incorporate Chapter 8, The Slice Sampler, into the Gibbs-sampler chapters as a special case, I can see the pedagogical rational for introducing it first 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.

With perhaps the exception of Chapter 12, Diagnosing Convergence, the rest of the book is devoted to materials that are less palatable for the first-timers. Indeed, initially 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 authors' skillful presentations have 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 that 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 yet nothing is 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 ten chapters; for recently developed recipes one can consult the Handbook of Markov Chain Monte Carlo (Brooks et al., 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 has been far more relaxing than I initially expected, at least intellectually!

Acknowledgements

I thank Radu Craiu and Stephen Finch for very helpful comments and editing. (Xiao-Li Meng)

Further references

Meyn, S.P. and Tweedie, R.L. (1993). Markov Chains and Stochastic Stability Springer-Verlag, New York.

Brooks, S., Gelman, A., Jones, G. and Meng, X.L. (2011) Handbook of Markov Chain Monte Carlo CRC Press, Boca Raton.