Book reviews, Winter 2011

Christian Robert

Book reviews, Winter 2012

Christian P. Robert
Université Paris-Dauphine, CEREMADE, IUF, and CREST
xian@ceremade.dauphine.fr
http://xianblog.wordpress.com

January 5, 2012

Abstract

This note is made of four book reviews of Brooks et al. (2011), Karian and Dudewicz (2011), McGrayne (2010), and Ziliak and McCloskey (2008), respectively. They are scheduled to appear in the next issue of *CHANCE*.

the theory that would not die, by Sharon Bertsch McGrayne

- **Hardcover:** 320+xiv pages
- **Publisher:** Yale University Press (first edition, May 2011)
- **Language:** English
- **ISBN-10:** 0300169698

A few days ago prior to reading her book and writing this review, I had lunch with the author of *the theory that would not die*, Sharon McGrayne, in a Parisian café and we had a wonderful chat about why she wrote the book and on the people she met during its completion. Among others, she mentioned the considerable support provided by Dennis Lindley, Persi Diaconis, and Bernard Bru. This conversation also acted as an introduction to the interview published in this issue of CHANCE. (The fact that I had not fully read the book before was due to delays in the delivery, presumably linked to the fact that the publisher, Yale University Press, had not forecasted the phenomenal success of the book and thus failed to scale the reprints accordingly!)

My reaction to the book is one of enthusiasm and glee! It indeed tells of the story and of the stories of Bayesian statistics and of Bayesians in a most entertaining if unmathematic manner. There will be some who will object to such a personification of science, which should be (much) more than the sum of the characters who contributed to it. Or even to the need to mention those characters, once the concepts they uncovered were incorporated within the theory. Overall, I share the materialist belief that those concepts are existing per se and thus would have been found, sooner or later, by A or X... However, and somehow paradoxically, I also support the perspective that, since (Bayesian) statistical science is as much philosophy as it is mathematics and computer-science, the components that led to its current state were contributed by individuals, for whom the path to those components mattered. Which is why I find the title particularly clever and, as Peter Müller pointed out, much more to the point than a sentence explicitly involving Bayes' Theorem.

While the book inevitably starts with the (patchy) story of Thomas Bayes's life, incl. his passage at Edinburgh University, and a nice non-mathematical description of his ball experiment, the next chapter is about "the man who did everything", Pierre-Simon (de) Laplace himself, for whom the author and myself share the same admiration. How Laplace attacked the issue of astronomical errors is brilliantly depicted, rooting the man within statistics and explaining why he would soon move to the "probability of causes". And rediscover plus generalise Bayes' theorem. That his (admittedly unpleasant!) thirst for honours and official positions would later cast disrepute on his scientific worth is difficult to fathom in restrospect. The next chapter is about the dark ages of [not yet] Bayesian statistics and I particularly liked the links with the French army, discovering there that the great Henri Poincaré testified at Dreyfus trial using a Bayesian argument, that Bertillon had completely missed the probabilistic point, and that the military judges were then all aware of Bayes' theorem, thanks to Bertrand's probability book being used at Ecole Polytechnique! (The last point actually was less of a surprise, given that I had collected some documents about the involvement of late 19th/early 20th century French artillery officers in the development of Bayesian techniques, Edmond Lhostes and Maurice Dumas, in connection with Lyle Broemeling's Biometrika study, 2003.) The description of the fights between Fisher and Bayesians and non-Bayesians alike is as always both entertaining and sad. Sad also is the fact that Jeffreys' (1939) masterpiece got so little recognition at the time. (While I knew about Fisher's unreasonable stand on smoking, going as far as defending the assumption that "lung cancer might cause smoking"(!), the Bayesian analysis of Jerome Cornfield was unknown to me. And quite fascinating.) The figure of Fisher actually permeates the whole book (and the one next reviewed!), as a negative and bullying figure preventing further developments of early Bayesian statistics, but also as an ambivalent anti-Bayesian who eventually tried to create his own brand of Bayesian statistics in the format of fiducial statistics (Seidenfeld, 1992).

> *"...and then there was the ghastly de Gaulle." D. Lindley*

The following part of *the theory that would not die* is about Bayesian contributions to the (second World) war, at least from the Allied side. Again, I knew most of Alan Turing' involvment in Bletchley Park's Enigma, however the story is well-told and, as in previous occasions, I cannot but be moved by the absurd waste of such a superb intellect by a blind administrative machine. The role of Albert Madansky in the assessment of the [lack of] safety of nuclear weapons is also well-described, stressing the inevitability of a Bayesian assessment of a one-time event that had [thankfully] not yet happened. The above quote from Dennis Lindley is the conclusion of his argument on why Bayesian statistics were not called Laplacean; I would suggest that the French post-war attraction for abstract statistics in the wake of Bourbaki also did a lot against this recognition in addition to de Gaulle's isolationism and ghastliness (or maybe they were one and the same thing). The involvement of John Tukey into military research was also a novelty for me, but not so much as his use of Bayesian [small area] methods for NBC election night previsions. The conclusion of Chapter 14 on why Tukey felt the need to distance himself from Bayesianism is quite compelling. Maybe paradoxically, I ended up appreciating Chapter 15 even more for the part about the search for a missing H-bomb near Palomares, Spain, as it exposes the plusses a Bayesian analysis would have brought.

> *"There are many classes of problems where Bayesian analyses are reasonable, mainly classes with which I have little acquaintance." J. Tukey*

When approaching near recent times and to contemporaries, Sharon McGrayne gives a very detailed coverage of the coming-of-age of Bayesians like Jimmy Savage and Dennis Lindley, as well as the impact of Stein's paradox (a personal epiphany!), along with the important impact of Howard Raiffa and Robert Schlaifer, both on business schools and on modelling prior beliefs [via conjugate priors]. I did not know anything about their scientific careers, but *Applied Statistical Decision Theory* (1961) is a beautiful book that prefigured both DeGroot's (1970) and Berger's (1985). (As an aside, I was amused by Raiffa using Bayesian techniques for horse betting based on race bettors, as I had vaguely played with the idea during my spare if compulsory time in the French Navy!) Similarly, while I had read detailed scientific accounts of Frederick Mosteller's and David Wallace's superb Federalist Papers study, they were only names to me. Chapter 12 mostly remedied this lack of mine's.

"We are just starting." P. Diaconis

The final part, entitled "Eureka!", is about the computer revolution we witnessed in the 1980s, culminating with the (re)discovery of MCMC methods we covered in our own "history". Because it contains stories that are closer and closer to today's time, it inevitably crumbles into shorter and shorter accounts. However, *the theory that would not die* conveys the essential message that Bayes rule had become operational, with its own computer language and objects like graphical models and Bayesian networks that could tackle huge amounts of data and real-time constraints. And used by companies like Microsoft and Google. The final pages mention neurological experiments on how the brain operates in a Bayesian-like way (a direction much followed by neurosciences).

In conclusion, I highly enjoyed reading through *the theory that would not die*. And I am sure most of my Bayesian colleagues will as well. Being Bayesians, they will compare the contents with their subjective priors about Bayesian history, but will in the end update those profitably. (The most obvious missing part is in my opinion the absence of E.T Jaynes and the MaxEnt community, which would deserve a chapter on its own.) As an insider, I have little idea on how the book would be perceived by the layman: it does not contain any formula apart from [the discrete] Bayes rule at some point, so everyone can read through. The current success of *the theory that would not die* shows that it reaches much further than academic circles. It may be that the general public does not necessarily grasp the ultimate difference between frequentist and Bayesians, or between Fisherians and Neyman-Pearsonians, as to why p-values should not be used (see also the next review). However *the theory that would not die* goes over all the elements that explain these differences. In particular, the parts about single events are quite illuminating on the specificities of the Bayesian approach. I will certainly [more than] recommend it to all of my graduate students.

Further references

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. Springer-Verlag, New York.

Broemeling, L. and Broemeling, A. (2003). Studies in the history of probability and statistics XLVIII The Bayesian contributions of Ernest Lhoste. *Biometrika*, **90** 728–731.

DeGroot, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

Jeffreys, H. (1939). *Theory of Probability*. 1st ed. The Clarendon Press, Oxford.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. MIT Press, Cambridge, Mass.

Seidenfeld, T. (1992). R.A. Fisher's fiducial argument and Bayes' theorem. *Statist. Science*, **7** 358–368.

The cult of significance, by Stephen Ziliak and Deirdre McCloskey

- **Hardcover:** 322+xxiv pages
- **Publisher:** The University of Michigan Press (First edition, Feb. 2008;
- **Language:** English
- **ISBN-13:** 978-0472050079

The book, written by economists Stephen Ziliak and Deirdre McCloskey, has a theme bound to attract both Bayesians and all those puzzled by the absolute and automatic faith in significance tests exhibited in many applied papers. The main argument of the authors is indeed that an overwhelming majority of papers involved in data analysis stop at rejecting variables ("coefficients") on the sole and unsupported basis of non-significance at the 5% level. Hence the subtitle: *How the standard error costs us jobs, justice, and lives* This is an argument I completely agree with, however, the aggressive style of the book ended putting me off as early as the first chapter! Obviously, I could have let both the matter and the book go, however I feel the book may in the end do a disservice to a valid issue and I thus endeavour to explain why through this review.

Significance testing

The category of significance tests addressed (and attacked) by *The cult of significance* is the point null single hypothesis of the form "is θ equal to zero?" commonly found in regression diagnoses. For instance, the first output of `demo(lm.glm)` in R (R Development Core Team, 2006) [a "canned regression package taking over the mind of the scientist"?, page 69] leads to

```
> summary(lm(weight~group-1))

Call:
lm(formula = weight ~ group - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0710 -0.4938  0.0685  0.2462  1.3690

Coefficients:
         Estimate Std. Error t value Pr(>|t|)
groupCtl   5.0320     0.2202   22.85 9.55e-15 ***
groupTrt   4.6610     0.2202   21.16 3.62e-14 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.6964 on 18 degrees of freedom
Multiple R-squared: 0.9818,     Adjusted R-squared: 0.9798
F-statistic: 485.1 on 2 and 18 DF,  p-value: < 2.2e-16
```

with the starred p-values `Pr(>|t|)` indicating whether or not the corresponding coefficients are "significant". (In the current example, they are both deemed to be significant.) Accepting or rejecting covariates on the sole basis of those stars ("asterisk econometrics", page 70) or of those p-values is highly reductive in that (a) they only consider the exact point null hypotheses, $H_0:\ \beta_i = 0$, when the exact nullity is so rarely at stake as to be a non-problem; (b) they do not account for the alternative hypothesis, nor for the relative likelihoods of both hypotheses (accepting the null hypothesis means the data is compatible with the null hypothesis, not that the null hypothesis is "true"); (c) they therefore fail to compare the predictive abilities of both representations (or models) and the relative variability of the estimates under both models; and (d) even within the null hypothesis referential, they fail to account for multiple testing.

> *"Advanced empirical economics, which we've endured, taught, and written about for years, has become an exercise in hypothesis testing, and is broken. We're saying the brokenness extends to many other quantitative sciences." (page xviii)*

The first chapters hardly contain any scientific argument, but rather imprecations against those blindly using significance tests. Rather than explaining in simple terms and with a few mathematical symbols [carefully avoided throughout the book] what the problem is with significance tests, Ziliak and McCloskey start with the assumption that the reader knows what tests are or, worse, that the reader does not even need to know. While the insistence on thinking about the (decisional) impact of a significant or insignificant coefficient/parameter in terms of the problem at hand is more than commendable, the alternative put forward by the authors remains quite vague, like “size matters”, “how big is big?”, and so on. They mention Bayesian statistics a few time, along with quotes of Jeffreys (1939) and Zellner (1984), but never get into the details of their perspective on model assessment. (In fact, the repeated call on determining how important the effect is seems to lead to some sort of prior on the alternative to the null.) It would have been fairly easy to pick one of the terrible examples ridiculed by Ziliak and McCloskey and to show the reader what a decent statistical analysis could have produced without more statistical sophistication than the one required by t-tests. Instead, the authors conducted a massive (and rather subjective) study of the *American Economic Review* for the 1980's with regard to the worth of all [statistical] significance studies used in all papers published in the journal, then repeated the analysis for the 1990's, and those studies constitute the core of their argument. (Following chapters reproduce the same type of analysis in other fields like epidemiology and psychometrics.)

Power in testing

The *power* of a testing procedure used to test the null hypothesis H_0 against the alternative hypothesis H_a is the probability of rejecting the null hypothesis under the alternative model. It therefore usually depends on an unknown parameter. For instance, if $H_0: \ \theta = 0$ and $H_a: \ \theta \neq 0$, the power is a function of θ. For a specific type of hypotheses and of sampling distributions, Neyman and Pearson were able to derive *most powerful tests* that, for a give Type I error (i.e., the error under the null hypothesis) maximise the power uniformly over all values of the parameter θ. This, however, is a very special occurrence and, in realistic situations, like other risk functions, power functions corresponding to two testing procedures will not get uniformly ordered for all values of the parameter. In such cases, the standard (if highly debatable) practice is to replace the unknown parameter by a (plug-in) estimate, resulting in the so-called ‘observed power’. As debated in Hoenig and Heisey (2001), high observed power when the null hypothesis cannot be significantly rejected should not be thought as a support for the null hypothesis. This seems to be the reasoning advocated by the authors.

> *"Fisher realized that acknowledging power and loss function would kill the unadorned significance testing he advocated and fought to the end, and successfully, against them." (page 144)*

Ziliak and McCloskey somehow surprisingly focus on the arch-villain Ronald Fisher while leaving Neyman and Pearson (mostly) free from their attacks. (And turning Gosset into the good fellow, supposed to be "hardly remembered nowadays" [p.3], while being dubbed a "lifelong Bayesian" [p.152].) I write "surprisingly" because Fisher did not advise as much the use of a fixed significance level (even though he indeed considered 5% as a convenient bound) as the use of the p-value *per se*, while Neyman and Pearson introduced fixed 5% significance levels as an essential part of their testing apparatus. (See Berger, 2003 and Denis, 2004.) Not a surprising choice when considering the unpleasant personality of Fisher, of course! (Another illustration of the focussed attack: "Fisherians do not literally conduct experiments. The brewer did." [p.27] conveniently omitting Fisher's career at Rothamsted experimental station.) On the opposite, the twined fathers of significance testing seem to escape the wrath of Ziliak and McCloskey due to their use of a loss function. Or maybe for defining a precise alternative. While I completely agree that loss functions should be used to decide about models (or predictives), the loss function imagined by Neyman and Pearson is simply too mechanistic to make any sense to a decision analyst. Or even to a statistician.

> *"Significance unfortunately is a useful mean towards personal ends in the advance of science, status and widely distributed publications, a big laboratory, a staff of research assistants, a reduction in teaching load, a better salary, the finer wines of Bordeaux (...) In a narrow and cynical sense statistical significance is the way to achieve these." (page 32)*

Once again, I find it quite sad that a book that addresses such an important issue let aggressiveness ruin its purpose. To the uneducated reader, it sounds too much like a crusade against an establishment to be convincing to neophytes and to be taken as a serious warning. (I wonder in fact what is the intended readership of this book, given that it requires some statistical numeracy, but not "too much" to be open-minded about statistical tests!) Bullying certainly does not help in making one's case more clearly understood, especially in scientific matters: even though letting mere significance tests at standard levels rule the analysis of a statistical model is a sign of intellectual laziness, or of innumeracy, accusing its perpetrators of intentional harm and cynicism as in the above quote does not feel adequate.

While I fully agree that users of statistical methods should not let SAS (or any other software) write the statistical part of their research paper for them but, instead, think about the indications provided by such outputs in terms of the theory and concepts behind their model(s). Interestingly, Ziliak and McCloskey mention for instance the use of simulation and pseudo-data to reproduce the performance of those tests under the assumed model and to calibrate the meaning of tools like p-values. A worthwhile and positive recommendation in an otherwise radically negative and possibly counter-productive book.

> *"Adam Smith, who is much more than an economist, noted in 1759 that hatred, resentment, and indignation against bad behaviour serve, of course, a social purpose (...) "Yet there is still something disagreeable in the passions themselves"." (page 55)*

The first example Ziliak and McCloskey use to make their point falls quite far from the mark: in Chapter 1, discussing the impact of two diet pills A and B with means 20 and 5 and standard deviations 5 and 1/4, respectively, they conclude that B gives a smaller p-value for the test of whether the pill has no effect. Because 20/10=2 and 5/(1/2)=10. Hence demonstrating how p-values can go wrong. There are two misleading issues there: first, the diets are compared in terms of mean effect, so outside statistics. Second, running a t-test of nullity of the mean is not meaningful in this case. What imports is whether or not a diet is more efficient than the other. Assuming a normal distribution, we have here

$$P(A > B) = P(X > -15/\sqrt{25 + 1/16}) = \Phi(2.996) = 0.999\,,$$

which sounds like a rather good argument in favour of diet pill A. (Of course, this is under the normal assumption and all that, which can be criticised and assessed.) The surprising thing is that Ziliak and McCloskey correctly criticise a similar error made in the New Jersey vs. Pennsylvania minimum wage study (Chapter 9, pp.101-103).

> *"Fisher-significance is a manly sounding answer, though false. And one can see in the dichotomy of hard and soft a gendered worry, too. The worry may induce some men to cling to Significance Only (...) Around 1950, at the peak of gender anxiety among middle-class men in the United States, nothing could be worse than to call a man soft." (pages 140-141)*

The above quote is completely unrelated to the issue and illustrates the level of irrational non-academic arguments reached at times by *the cult of* *statistical significance*. (The authors also blame the massacre of whales and the torturing of lambs, p. 39, on t-tests!) Just as laughable is the characterisation of statistics as the "bourgeois cousin" of probability theory (p.195) at a time where both fields did not truly exist and were clearly mixed in most researchers' mind (as shown by the titles of Keynes' and Jeffreys' books). In addition, this dismissive "bourgeois" adjective reproduces the scorn exhibited by some pure mathematics circles for statistics, because of its applied inclinations. (Evidently such a sociological qualificative does not bring any academic argument to the validity of one's viewpoint.)

> *"The overall combative style rapidly becomes grating." David Aldous, Amazon review (2008)*

As a final note, let me point out that this book got published in 2008, hence already received a lot of reviews, both in economics and in statistics, some of which are available on the authors' webpages. At the junction between philosophy of science and econometrics, Aris Spanos reviewed *the cult of statistical significance* from the Fisher-Neyman-Pearson fusion perspective developed in *Error and Inference*, edited by Mayo and Spanos (2010). David Aldous also wrote a convincing and balanced short review on Amazon about the book.

Further references

Berger, J. (2003). Could Fisher, Jeffreys and Neyman have agreed on testing? *Statistical Science*, **18** 1–12. (with discussion).

Denis, D. (2004). The modern hypothesis testing hybrid: R.A. Fisher's fading influence. *J. Soc. Française Statist.*, **145** 5–26. (with discussion).

Hoenig, J. and Heisey, D. (2001). The abuse of power. *The American Statistician*, **55** 19–24.

Jeffreys, H. (1939). *Theory of Probability.* 1st ed. The Clarendon Press, Oxford.

Mayo, D. and Spanos, A. (2010). *Error and inference: recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science.* Cambridge University Press.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Zellner, A. (1984). *Basic issues in econometrics.* University of Chicago Press, Chicago.

Handbook of Markov chain Monte Carlo, edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng

- **Hardcover:** 619 pages
- **Publisher:** Chapman and Hall/CRC Press (first edition, May 2011)
- **Language:** English
- **ISBN-10:** 1420079417

At JSM this summer, John Kimmel from Chapman and Hall/CRC Press gave me a copy of the *Handbook of Markov chain Monte Carlo*, as I had not received an author's copy. This handbook is edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, all top jedis of the MCMC galaxy. (Note that there was an MCMC "sidebar" in my Book Reviews column in the previous issue of CHANCE.) Of course, authors and editors being friends of mine, the reader may worry about the objectivity of this assessment; she or he needs not, the quality of the contents is clearly there and the book appears as a worthy successor to the tremendous *Markov chain Monte Carlo in Practice* by Wally Gilks, Sylvia Richardson and David Spiegelhalter (1996). (As an author, I can attest to the involvement of the current editors from the many rounds of reviews we exchanged about our MCMC history chapter!) The style of the chapters is rather homogeneous and there are a few R codes here and there. So, while I will still stick to our book (Robert and Casella, 2004) for teaching MCMC to my graduate students this semester, I think the *Handbook* can well be used at a teaching level as well as a reference on the state-of-the-art MCMC technology.

Perfect sampling

When running an MCMC algorithm, the main worry is to know *when to stop*! The algorithm is indeed based on the mathematical theorem that, when t goes to infinity, the current value of the simulated Markov chain, x^t, will be distributed from the target distribution f. In practice, the usual if lazy attitude is to set a fixed number of iterations, $10,000$ say, and hope for the best. There actually exist a whole range of techniques to check during or after simulation whether or not the chain "has converged" (or at least if some feature of the convergence can be observed). Those are covered for instance in our R book (Robert and Casella, 2010, Chap. 8).

As a formal alternative, there also exist techniques that guarantee an output from f when based on a MCMC algorithm, possibly at a considerable computing expense. Such techniques are called *perfect* (or exact) *sampling* or yet *coupling from the past* and were popularized by Propp and Wilson (1996). and Kendall and Møller (1999). Put into words, the principle of perfect sampling is to start the Markov chain $(x^t)_t$ at time $-\infty$ so that x^0 is in the stationary distribution, f, rather than starting it at time $t = 0$ and having to wait till time $+\infty$. While this sounds like a useless trick, the implementation of the principle is to start back enough in time to cancel the impact of the starting value by time $t = 0$. Once again, this sounds like another trick in that all possible starting values cannot be examined at once. However, there exist many ways to reduce the number of starting values to a finite set of "extreme" values. David Wilson operated a webpage on the topic till 2004[1] and Jeff Rosenthal wrote a nice applet to illustrate the principle for a random walk on a finite set, using two "extreme" values.[2] However, while the perspective of turning an MCMC sampler into an exact sampler is conceptually fascinating and did attract many MCMC researchers, the implementation of the perfect sampling idea is most often untractable for realistic problems, which explains why it is not more widespread.

I will not go in details over all the chapters (some are available online). The first half of the book covers MCMC methodology with a beautiful and lively first chapter by Charlie Geyer that manages to highlight the essentials of MCMC in a very coherent way while also explaining very very clearly the four fundamentals advances contained in Peter Green's (1995) reversible jump paper. (I figure it would seem like base-jumping to someone who had never heard of MCMC! In the literal sense of jumping from a cliff with 5 seconds to reach the ground!) Terrific chapter! While it would have been equally terrific to read the expected chapter on reversible jump by Peter Green and David Hastie, Yanan Fan and Scott Sisson survey reversible jump in proper details in Chapter 3, esp. convergence assessment for RJMCM. Then, the next chapter about optimal proposal distributions and adaptive MCMC is from Jeff Rosenthal, with his usual pedagogical qualities (incl. great FAQ sections!). The chapter about MCMC using Hamiltonian dynamics is also from Toronto, being written by Radford Neal, and it is a huge chapter, full of details and ideas about Hamiltonian MCMC, that should prove very profitable to all readers. (And a good prequel to Girolami and Calderhead's Read Paper in 2011.)

[2] Site: http://research.microsoft.com/en-us/um/people/dbwilson/exact

[2] Site: http://probability.ca/jeff/java/cftp.html

Approximate Bayesian Computation (ABC)

ABC methods were introduced by Tavaré et al. (1997) as a manageable way to handle ungainly likelihood functions, namely in setting where regular MCMC methods (see past issue) do not apply because $f(\mathbf{y}|\theta)$ cannot be computed. The idea at the core of ABC is to simulate from the prior distribution $\pi(\theta)$ until a simulated pseudo-dataset is similar enough to the observed dataset $\mathbf{y}$. The similarity is defined in terms of a distance between summaries of the data, $\eta(\mathbf{y})$, and of a maximum tolerance ϵ over this distance. The algorithm runs as follows:

Algorithm 1 ABC algorithm

```
for i = 1 to N do
  repeat
    generate θ' from the prior distribution π(·)
    generate z from the likelihood f(·|θ')
  until ρ{η(z), η(y)} ≤ ϵ
  set θ_i = θ'
end for
```

It provides an approximation to the posterior distribution $\pi(\theta|\eta(\mathbf{y}))$ and thus is not as informative as the true posterior $\pi(\theta|\mathbf{y})$. However, in complex settings such as phylogenies, there is no available alternative for conducting inference about the parameters of the model and one has to agree on this departure from exact Bayesian inference. The method is therefore quite popular in population genetics (Cornuet et al., 2008), but also in financial and extreme modelling.

Both following chapters are about convergence assessments, by Andrew Gelman and Kenneth Shirley, and by James Flegal and Galin Jones. (Both give relevant advices. As stressed in our R book *Introducing Monte Carlo Methods with R*, 2010, I particularly like the idea of Flegal and Jones, 2008, to validate a bootstrap approach to confidence evaluation!) The next two chapters are covering perfect sampling, by Radu Craiu and Xiao-Li Meng, and by Mark Huber. (Perfect stuff, even though I got disillusioned over the years about the range of this fascinating use of MCMC outputs. Mark's spatial processes are certainly the most convincing domain of application.) Jim Hobert wrote a chapter on data augmentation algorithm, full of fine details about the convergence of this special case of Gibbs sampling, which illustrates very well the current thoughts on convergence assessment. Charlie Geyer has a short chapter on importance sampling, simulated tempering and umbrella sampling, with an application to the approximation of Bayes factors, while Scott Sisson and Yanan Fan wrote the chapter on ABC. (Two interesting sentences from this chapter are that "model comparison through likelihood-free posteriors with a fixed vector of summary statistics will ultimately compare distortions of those models which are overly simplified wrt the true data-generating process. This remains true even when using sufficient statistics and for $\epsilon \to 0$." (p.329) and "While [using likelihood-free inference for model selection purposes] is a natural extension of inference for individual models, the analysis in Section 12.4.4 urges caution and suggests that further research is needed into the effect of the likelihood-free approximation (...) on the marginal likelihoods upon which model comparison is based" (p.333), as our 2011 PNAS paper brings some light on both questionings.) The second half of the book is more topical, with applications of MCMC in Genetics, Physics, Ecology, MRI data, Astronomy, however it also contains methodological directions, like the chapter written by Paul Fearnhead on MCMC for state space models.

Further references

Cornuet, J.-M., Santos, F., Beaumont, M., Robert, C., Marin, J.-M., Balding, D., Guillemaud, T. and Estoup, A. (2008). Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics*, **24** 2713–2719.

Flegal, J., Haran, M. and Jones, G. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Science*, **23** 250–260.

Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). *Markov chain Monte Carlo in Practice*. Chapman and Hall, New York.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. Royal Statist. Society Series B*, **73** 123–214.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, **82** 711–732.

Kendall, W. and Møller, J. (1999). Perfect Metropolis–Hastings simulation of locally stable point processes. Tech. Rep. R-99-2001, Department of Mathematical Sciences, Aalborg Univ.

Propp, J. and Wilson, D. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9** 223–252.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. 2nd ed. Springer-Verlag, New York.

Robert, C. and Casella, G. (2010). A history of Markov chain Monte Carlo-subjective recollections from incomplete data. In *Handbook of Markov Chain Monte Carlo: Methods and Applications* (S. Brooks, A. Gelman, X. Meng and G. Jones, eds.). Chapman and Hall, New York. ArXiv0808.2902.

Robert, C., Cornuet, J.-M., Marin, J.-M. and Pillai, N. (2011). Lack of confidence in ABC model choice. *Proc. National Academy of Sciences*, **108** 15112–15117.

Tavaré, S., Balding, D., Griffith, R. and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, **145** 505–518.

Handbook of fitting statistical distributions with R, by Z. Karian and E.J. Dudewicz

- **Hardcover:** xlv+1672 pages+1 CD-ROM (6 pounds in weight, costing 80 pounds)
- **Publisher:** CRC Press, Taylor and Francis Group, Chapman & Hall, Boca Raton (first edition, Oct. 2010)
- **Language:** English
- **ISBN-13:** 978-1584887119

Yet another handbook?! When I received this book last July, I was obviously impressed by its size (around 1700 pages and 3 kilos...). From briefly glancing at the table of contents and at the list of standard distributions appearing as subsections of the first chapters, I thought that the authors were covering different estimation/fitting techniques for most of the standard distributions. After taking a closer look at the book, I think the cover is misleading in several aspects: this is not a handbook (a.k.a. a reference book), it does not cover standard statistical distributions, the R input is marginal, and the authors only wrote part of the book, since about half of the chapters are written by other authors, while recycling an earlier version of the book (Karian and Dudewicz, 2000).

> *"The system we develop in this book has its origins in the one-parameter lambda distribution proposed by John Tukey." Z. Karian and E.J. Dudewicz (page 3)*

Therefore I am quite glad I left *Handbook of fitting statistical distributions with R* in my office rather than dragging those three kilos along during my summer vacations, as I originally planned! First, the book indeed does not aim at fitting standard distributions but instead at promoting a class of quantile distributions, first introduced by Ramberg and Schmeiser (1974), the generalised lambda distributions (GLDs), whose quantile function is a location-scale transform of

$$Q(y|\lambda_3, \lambda_4) = F_X^{-1}(y) = y^{\lambda_3} - (1-y)^{\lambda_4}$$

(under the constraint on the parameters that the above function of y is non-decreasing for a positive scale and non-increasing otherwise) and that the authors have been advocating for a long while. There is nothing wrong *per se* with those quantile distributions, but neither is there a particular reason to prefer them over the standard parametric distributions! Overall, I am quite wary of one-fits-all distributions, especially when they only depend on four parameters and mix finite with infinite support distributions. The lack of natural motivations for the above is enough to make fitting with those distributions not particularly compelling. Karian and Dudewicz spend an awful lot of space on numerical experiments backing their argument that the generalised lambda distributions approximate reasonably well (in the L_1 and L_2 norm senses, as it does not work for stricter norms) "all standard" distributions, but it does not explain why the substitution would be of such capital interest. Furthermore, the estimation of the parameters (i.e. the "fitting" in *fitting statistical distributions*) is not straightforward. While the book presents the density of the generalised lambda distributions as available in closed form (Theorem 1.2.2), namely (omitting the location-scale parameters),

$$f(x|\lambda_3, \lambda_4) = \frac{1}{\lambda_3 F_X(x|\lambda_3, \lambda_4)^{\lambda_3 - 1} + \lambda_4\{1 - F_X(x|\lambda_3, \lambda_4)\}^{\lambda_4 - 1}},$$

it fails to state explicitly that the cdf

$$F_X(x|\lambda_3, \lambda_4) = Q^{-1}(x|\lambda_3, \lambda_4)$$

itself is not available in closed form. Therefore, neither likelihood estimation nor Bayesian inference seem easily implementable for those distributions. (Actually, a mention is made of maximum likelihood estimators for the first four empirical moments in the second chapter, but it is alas mistaken, confusing the renormalisation of those moments used in the normal model with genuine maximum likelihood estimation.) Obviously, given that quantile distributions are easy to simulate, ABC would be a manageable tool for handling Bayesian inference on GLDs... The book focus instead on moment and percentile estimators as the central estimation tool, with no clear message on which side to prefer (see, e.g., Section 5.5).

Quantile distribution and simulation

One appeal of quantile distributions like the one covered in this book is that they are easy to simulate. In fact, when the quantile function $Q(\cdot|\theta)$ is available in closed form, simulation from the associated distribution can be done by the mere transform

$$X = Q(U|\theta)\,, \qquad U \sim \mathcal{U}(0,1)\,,$$

of a uniform generation. This means that simulation-based inference methods like ABC (see previous sidebar), parametric bootstrap, or indirect inference, can easily be implemented for those distributions.

A chapter (by S. Su) covers the case of mixtures of GLDs, whose appeal is similarly lost on me. My major issue with using such distributions in mixture setting is that some components may have a finite support, which makes the use of score equations awkward and of Kullback-Leibler divergences to normal mixtures fraught with danger (since those divergence may then be infinite). The estimation method switches to maximum likelihood estimation, as presumably the moment method gets too ungainly. However, I fail to see how maximum likelihood is implemented: I checked the original paper by Su (2007), documenting the related `GLDEX` R function, but the approach is very approximate in that the true percentiles are replaced with pluggin (and fixed, i.e. non-iterative) values (again omitting the location-scale parameters)

$$\hat{u}_i = F(x_i|\hat{\lambda}_3, \hat{\lambda}_4) \qquad i = 1, ..., n$$

in the likelihood function

$$\prod_{i=1}^{n} \frac{1}{\lambda_3 \hat{u}_i^{\lambda_3 - 1} + \lambda_4 \{1 - \hat{u}_i\}^{\lambda_4 - 1}}$$

A further chapter is dedicated to the generalised beta distribution, which simply is a location-scale transform of the regular beta distribution (even though it is called the extended GLD for no discernible reason). Again, I have nothing for or against this family (except maybe that using a bounded support distribution to approximate infinite support distributions could induce potential drawbacks...) I simply cannot see the point in multiplying parametric families of distributions where there is no compelling property to do so. (Which is also why as an editor/aeditor/referee, I have always been ultra-conservative vis--vis papers introducing new families of distributions.)

The R side of the book (i.e. the "R" part in *fitting statistical distributions with R*) is not particularly appealing either: in the first chapters, i.e. in the first hundred pages, the only reference to R is the name of the R functions found on the attached CD-ROM to fit GLDs by the method of moments or of percentiles... The first detailed code is found on pages 305-309, but it is unfortunately a MATLAB code! (Same thing in several subsequent chapters.) Even though there is an R component to the book thanks to this CD-ROM, the authors could well be suspected of "surfing the R wave" of the *Use R!* and other "with R" collections. Indeed, my overall feeling is that they are mostly recycling their 2000 book *Fitting statistical distributions* into this R edition. (For instance, figures that are reproduced from the earlier book, incl. the cover, are not even produced with R. Most entries of the table of contents of "Fitting statistical distributions" are found in the table of contents of *Handbook of fitting statistical distributions with R*. The codes were then written in Maple and some Maple codes actually survive in the current version. Most of the novelty in this version is due to the inclusion of chapters written by additional authors.)

> *"It remains for a future research topic as to how to improve the generalised bootstrap to achieve a 95% confidence interval since 40% on average and 25%-55% still leaves room for improvement." W. Cai and E.J. Dudewicz (page 852)*

As in the 2000 edition, the "generalised bootstrap" method is argued as an improvement over the regular bootstrap, "fraught with danger of seriously inadequate results" (p.816), and as a mean to provide confidence assessments. This method, attributed to the authors in 1991, is actually a parametric bootstrap used in the context of the GLDs, where samples are generated from the fitted distribution and estimates of the variability of estimators of interest are obtained by a sheer Monte Carlo evaluation! (A repeated criticism of the bootstrap is its "inability to draw samples outside the range of the original dataset" (e.g., p.852). It is somehow ironical that the authors propose to use instead parameterised distributions whose support may be bounded.)

Among other negative features of the book, I want to mention the price ($150!!!), the glaring [for statisticians!] absence of confidence statements about the (moment and percentile) estimations (not to be confused with goodness-of-fit)—except for the much later chapter on generalised bootstrap—, the fact that the book contains more than 250 pages of tables—yes, printed tables as in Gosset's era!—including a page with a few hundred random numbers generated from a given distribution, the fact that the additional authors who wrote the contributed chapters are not mentioned elsewhere that in the front page of those chapters—not even in the table of contents—, and, to repeat the fact once more, the misleading use of the term *handbook* in the title, the way *Wiktionary* defines it.

Further references

Karian, Z. and Dudewicz, E. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods.* Chapman and Hall, New York.

Marin, J., Pudlo, P., Robert, C. and Ryder, R. (2011). Approximate Bayesian computational methods. *Statistics and Computing*, **21** 289–291.

Ramberg, J. and Schmeiser, B. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, **17** 78–82.

Su, S. (2007). Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R. *J. Statist. Software*, **21** 9.