



HAL
open science

Book reviews, Summer 2012

Christian Robert

► **To cite this version:**

Christian Robert. Book reviews, Summer 2012. CHANCE, 2012, 25 (3), pp.58-62.
10.1080/09332480.2012.726567 . hal-00687315

HAL Id: hal-00687315

<https://hal.science/hal-00687315v1>

Submitted on 12 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Book reviews, Summer 2012

Christian P. Robert^{1,2,3}

¹Université Paris-Dauphine, CEREMADE, ²IUF, and ³CREST
xian@ceremade.dauphine.fr

April 12, 2012

Abstract

This note is made of a review of the books by Cox and Donnelly (2011), Efron (2011), Everitt (2012), Belgio (2011), respectively. They are scheduled to appear in the next issue (25(3)) of *CHANCE*.

Principles of Applied Statistics by David R. Cox and Christl Donnelly

This book by David Cox and Christl Donnelly is an extensive if condensed coverage of most (all?) necessary steps and precautions one must go through when contemplating applied (i.e. actual!) statistics. As the authors write in their very first sentence of the book, “applied statistics is more than data analysis” (p.i); the title could thus have been “Principled Data Analysis”! Indeed, *Principles of Applied Statistics* reminds me of how much we (at least I) take ‘the model’ and ‘the data’ for granted when conducting statistical analyzes, by going through all the pre-data and post-data steps that lead to the “idealized” (p.188) data analysis. The contents of the book are intentionally simple, with hardly any mathematical aspect, but with a clinical attention to exhaustivity and clarity. For instance, even though I would have enjoyed more stress on probabilistic models as the basis for statistical inference, they only appear by the fourth chapter (out of ten) with error in variable models. The painstakingly careful coverage of the myriad of tiny but essential steps involved in a statistical analysis and the highlight of the numerous corresponding pitfalls was certainly illuminating to me. Just as the book refrains from mathematical digressions (“our emphasis is on the subject-matter, not on the statistical techniques as such” p.12), it refrains from engaging into detail and complex data stories. Instead, it uses little

grey boxes to convey the pertinent aspects of a given data analysis, referring to a paper for the full story. (I must admit this is frustrating at times, as one would like to read more!) The book reads very nicely and smoothly, and I must acknowledge I read most of it in trains, métros, and planes over a week.

‘A general principle, sounding superficial but difficult to implement, is that analyses should be as simple as possible, but not simpler.’ Cox and Donnelly (p.9)

To get into more details, *Principles of Applied Statistics* covers most purposes of statistical analyses (Chap. 1), design with some special emphasis (Chap. 2-3), which is not surprising given the record of the authors (and “not a moribund art form”!, p.51), measurement (Chap. 4), including the special case of latent variables and their role in model formulation, preliminary analysis (Chap. 5) by which the authors mean data screening and graphical pre-analysis, [at last!] models (Chap. 6-7), separated in model formulation [debating the nature of probability] and model choice, the later being somehow separated from the standard meaning of the term (covered in §8.4.5 and §8.4.6), formal [mathematical] inference (Chap. 8), handling in particular testing and multiple testing, interpretation (Chap. 9), i.e. post-processing, and a final epilogue (Chap. 10). The intended readership of the book is rather broad, from practitioners to students, although both categories do require a good dose of maturity to fully appreciate the book, to teachers, to scientists designing experiments with a statistical mind. It may be deemed too philosophical by some, too allusive by others, but I think it constitutes a magnificent testimony to the depth and to the spectrum of our field.

‘Of course, all choices are to some extent provisional.’ Cox and Donnelly (p.130)

As a personal aside, I clearly appreciated the illustration using capture-recapture models (p.36) with a remark of the impact of toe-clipping on frogs, as it reminded me of a similar way of marking lizards when my (then) PhD student Jérôme Dupuis (1995 was working on a corresponding capture-recapture dataset from Southern France. On the opposite, while John Snow’s story [of using maps to explain the cause of cholera] is alluring, and his map makes for a great cover (!), I am less convinced it is particularly relevant within this book, given that Snow’s scientific inference was conducted without the map, later used to convince local authorities.

‘The word Bayesian, however, became more widely used, sometimes representing a regression to the older usage of flat prior distributions supposedly representing initial ignorance, sometimes meaning models in which the parameters of interest are regarded as random variables and occasionally meaning little more than that the laws of probability are somewhere invoked.’ Cox and Donnelly (p.144)

My main quibble with the book goes, most unsurprisingly!, with the processing of Bayesian analysis found in *Principles of Applied Statistics* (pp.143-144). Indeed, on the one hand, the method is mostly criticized over those two pages. On the other hand, it is the only method presented with this level of details, including historical background, which seems a bit superfluous for a treatise on applied statistics. The drawbacks mentioned are (p.144)

the weight of prior information or modelling as ‘evidence’;

the impact of ‘indifference or ignorance or reference priors’;

whether or not empirical Bayes modelling has been used to construct the prior;

whether or not the Bayesian approach is anything more than a ‘computationally convenient way of obtaining confidence intervals’.

The empirical Bayes perspective is the original one found in Robbins (1955) and seems to find grace in the authors’ eyes (“the most satisfactory formulation”, p.156). Contrary to MCMC methods, “a black box in that typically it is unclear which features of the data are driving the conclusions” (p.149)... A bit drastic an appreciation!

*‘If an issue can be addressed nonparametrically then it will often be better to tackle it parametrically; however, if it cannot be resolved non-parametrically then it is usually dangerous to resolve it parametrically.’
Cox and Donnelly (p.96)*

Apart from a more philosophical paragraph on the distinction between machine learning and statistical analysis in the final chapter, with the drawback of using neural nets and such as black-box methods (p.185), there is relatively little coverage of non-parametric models in the book, the choice of “parametric formulations” (p.96) being openly chosen. I can somehow understand this perspective for simpler settings, namely that non-parametric models offer little explanation of the production of the data. However, in

more complex models, non-parametric components often are a convenient way to evacuate burdensome nuisance parameters. Again, technical aspects are not the focus of *Principles of Applied Statistics* so this also explains why it does not dwell intently on non-parametric models.

‘A test of meaningfulness of a possible model for a data-generating process is whether it can be used directly to simulate data.’ Cox and Donnelly (p.104)

The above remark is quite interesting, especially when accounting for David Cox’ current appreciation of ABC techniques (see my vignette on ABC in the 24(4) issue of CHANCE). The impossibility to generate from a posited model as some found in econometrics precludes using ABC, but this does not necessarily mean the model should be excluded as unrealistic.

‘The overriding general principle is that there should be a seamless flow between statistical and subject-matter considerations.’ Cox and Donnelly (p.188)

As mentioned earlier, the last chapter brings a philosophical conclusion on what is (applied) statistics. It stresses the need for a careful and principled use of black-box methods so that they preserve a general framework and lead to explicit interpretations. Once again, a must-read for all statisticians!

Further reading

DUPUIS, J. (1995). Bayesian estimation of movement probabilities in open populations using hidden Markov chains. *Biometrika*, **82** 761–772.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. In *Proc. Third Berkeley Symp. Math. Statist. Probab.*, vol. 1. University of California Press.

Large-scale Inference by Brad Efron

- **Paperback:** 276 pages
- **Publisher:** Cambridge University Press and Institute of Mathematical Statistics
- **Year:** 2010

- **Language:** English
- **ISBN-13:** 978-0-5211-9249-1

Large-scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction by Brad Efron is the first IMS Monograph in this new series, coordinated by David Cox and published by Cambridge University Press. Since I read this book immediately after Cox' and Donnelly's *Principles of Applied Statistics*, reviewed above, I was thinking of drawing a parallel between the two books. However, while none of them can be classified as textbooks [even though Efron's contains exercises], they differ very much in their intended audience and in their purpose. As I wrote in the review of *Principles of Applied Statistics*, the book has an encompassing scope with the goal of covering all the methodological steps required by a statistical study. In *Large-scale Inference*, Efron focus on empirical Bayes methodology for large-scale inference, by which he mostly means multiple testing (rather than, say, data mining). As a result, the book is centered on mathematical statistics and is more technical. (Which does not mean it less of an exciting read!) The book was recently reviewed by both Michael Chernick and Jordi Prats for *Significance*. Akin to the previous reviewer, and unsurprisingly, I found the book nicely written, with a wealth of R (color!) graphs (the R programs and dataset are available on Brad Efron's home page).

'I have perhaps abused the "mono" in monograph by featuring methods from my own work of the past decade.' Brad Efron (p.xi)

Sadly, I cannot remember if I read my first Efron's paper via his 1977 introduction to the Stein phenomenon with Carl Morris in *Pour la Science* (the French translation of *Scientific American* I was reading at the time), following their 1975 seminal paper, or through his 1983 [still] *Pour la Science* paper with Persi Diaconis on computer intensive methods. (I would bet on the later though. And not only because it makes this Book Review tribune more homogeneous!) In any case, I certainly read a lot of Efron's papers on the Stein phenomenon during my PhD thesis and it was thus with great pleasure that I saw he introduced empirical Bayes notions through the Stein phenomenon (Chapter 1). It actually took me a while but I eventually (by page 90) realized that empirical Bayes was a proper subtitle to Large-Scale Inference in that the large samples were giving some weight to the validation of empirical Bayes analyses. In the sense of reducing the importance of a genuine Bayesian modelling (even though I do not see why this genuine

Bayesian modelling could not be implemented in the cases covered in the book).

Empirical Bayes principles Empirical Bayes methods can crudely be seen as the [information] poor man's Bayesian analysis! They start from a Bayesian modelling, for instance the parameterized prior

$$x \sim f(x|\theta), \quad \theta \sim \pi(\theta|\alpha)$$

and then, instead of setting α to a specific value or of assigning an hyperprior to this hyperparameter α , as in a regular or a hierarchical Bayes approach, the empirical Bayes paradigm consists in *estimating α from the data*. Hence the "empirical" label, using the data to build the "prior" instead of prior information or of objective Bayes principles (Robert, 2001). The reference model used for the estimation is the integrated likelihood (or conditional marginal)

$$m(x|\alpha) = \int f(x|\theta)\pi(\theta|\alpha) d\theta$$

which defines a distribution density indexed by α and thus allows for the use of any statistical estimation method (moments, maximum likelihood or even Bayesian!). A classical example is provided by the normal exchangeable sample: if

$$x_i \sim \mathcal{N}(\theta_i, \sigma^2) \quad \theta_i \sim \mathcal{N}(\mu, \tau^2) \quad i = 1, \dots, p$$

then, marginally,

$$x_i \sim \mathcal{N}(\mu, \tau^2 + \sigma^2)$$

and μ can be estimated by the empirical average of the observations, \bar{x} . The next step in an empirical Bayes analysis is to act as if α had not been estimated from the data and to conduct a regular Bayesian processing of the data with this estimated prior distribution. In the above normal example, this means estimating the θ_i 's by

$$\frac{\sigma^2 \bar{x} + \tau^2 x_i}{\sigma^2 + \tau^2}$$

with the characteristic shrinkage (to the average) property of the resulting estimator (Efron and Morris, 1975).

While using Bayesian tools, this technique is outside of the Bayesian paradigm for several reasons: (a) the prior depends on the data, hence it lacks foundational justifications; (b) the prior varies with the data, hence it lacks theoretical validations like Walk's complete class theorem; (c) the prior uses the

data once, hence the posterior uses the data twice (see the vignette about this “sin” in the previous issue); (d) the prior relies of an estimator, whose variability is not accounted for in the subsequent analysis (Morris, 1983). The original motivation for the approach (Robbins, 1955) was more non-parametric, however it gained popularity in the 70’s and 80’s both in conjunction with the Stein effect and as a practical mean of bypassing complex Bayesian computations. As illustrated by Efron’s book, it recently met with renewed interest in connection with multiple testing.

‘Large N isn’t infinity and empirical Bayes isn’t Bayes.’ Brad Efron (p.90)

The core of *Large-scale Inference* is multiple testing and the empirical Bayes justification/construction of Fdr’s (false discovery rates). Efron wrote more than a dozen papers on this topic, covered in the book and building on the groundbreaking and highly cited *Series B* 1995 paper by Benjamini and Hochberg. (In retrospect, it should have been a Read Paper and so was made a retrospective read paper by the Research Section of the RSS.) Fdr’s are essentially posterior probabilities and therefore open to empirical Bayes approximations when priors are not selected. Before reaching the concept of Fdr’s in Chapter 4, Efron goes over earlier procedures for removing multiple testing biases. As shown by a section title (‘Is FDR Control “Hypothesis Testing”?’’, p.58), one major point in the book is that an Fdr is more of an estimation procedure than a significance-testing object. (This is not a surprise from a Bayesian perspective since the posterior probability is an estimate as well.)

‘Scientific applications of single-test theory most often suppose, or hope for rejection of the null hypothesis (...) Large-scale studies are usually carried out with the expectation that most of the N cases will accept the null hypothesis.’ Brad Efron (p.89)

On the innovations proposed by Efron and described in *Large-scale Inference*, I particularly enjoyed the notions of local Fdrs in Chapter 5 (essentially pluggin posterior probabilities that a given observation stems from the null component of the mixture) and of the (Bayesian) improvement brought by empirical null estimation in Chapter 6 (‘not something one estimates in classical hypothesis testing’, p.97) and the explanation for the inaccuracy of the bootstrap (which ‘stems from a simpler cause’, p.139), but found

less crystal-clear the empirical evaluation of the accuracy of Fdr estimates (Chapter 7, ‘independence is only a dream’, p.113), maybe in relation with my early career inability to explain Morris’s (1983) correction for empirical Bayes confidence intervals (pp. 12-13). I also discovered the notion of enrichment in Chapter 9, with permutation tests resembling some low-key bootstrap, and multiclass models in Chapter 10, which appear as if they could benefit from a hierarchical Bayes perspective. The last chapter happily concludes with one of my preferred stories, namely the missing species problem (on which I hope to work this very Spring).

Further reading

CASELLA, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician*, **39**(2) 83-87.

EFRON, B. and MORRIS, C. (1975). Data analysis using Stein’s estimator and its generalizations. *J. American Statist. Assoc.*, **70** 311–319.

MORRIS, C. (1983). Parametric empirical Bayes inference: theory and applications. *J. American Statist. Assoc.*, **78** 47–65.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. In *Proc. Third Berkeley Symp. Math. Statist. Probab.*, vol. 1. University of California Press.

ROBERT, C. (2001). *The Bayesian Choice*, Springer-Verlag, New York.

A Whistle-Stop Tour of Statistics by Brian S. Everitt

- **Paperback:** xi+199 pages
- **Publisher:** CRC Press, Boca Raton FL
- **Year:** 2012
- **Language:** English
- **ISBN-13:** 978-1-4398-7748-7

‘The book is intended as a quick source of reference and as an aide-memoir [sic] for students taking A-level, undergraduate or postgraduate statistics courses.’ B. Everitt, page xi

In a consequent package I got from CRC Press and made of books proposed for review in CHANCE, there was this short book by Brian S. Everitt, *A Whistle-Stop Tour of Statistics*. Nice cover and fun title! The book is like an introductory undergraduate statistics course, except that it is much more terser and shorter, using 200 pages in A5 format with plenty of pictures. (It could have as well been called a primer or a guidebook.) The table of contents is as follows

Some basics and describing data

Probability

Estimation

Inference

Analysis of variance models

Linear regression models

Logistic regression and the generalized linear model

Survival analysis

Longitudinal data and their analysis

Multivariate data and their analysis

and thus covers most of the standard models analyzed in classical statistics textbooks. (Missing at least time series.)

There is nothing fundamentally wrong with the book, except that I cannot truly fathom its purpose, nor its readership. Once again, the book is way too short and terse to be used in an undergraduate course and even less for self-study. And it does not bring a new light on those standard topics when compared with most of introductory statistics books, being mostly traditional (even though it briefly mentions Bayesian inference on pp. 88-89). While the book is itself a summary of statistical methodology, it still finds room for a summary of the covered notions at the end of each chapter. Hence, I remain at this point utterly perplexed by the reason behind publishing *A Whistle-Stop Tour of Statistics*..!

Correlations between the Physical and Social Sciences by Valentine Belfiglio

- **Paperback:** 76 pages
- **Publisher:** University Press of America
- **Year:** 2011
- **Language:** English
- **ISBN-13:** 978-0-7618-5589-7

This is probably the most bizarre book I have received for review and I only include it in this column to illustrate how commercial editors seem to be so unconcerned by the contents and worth of scientific books. The cover of the book is enticing: a picture of young Albert Einstein. Its purpose is wide:

*‘The thesis of this monograph is that societies in general are governed by objective laws that have their roots in human nature. The task of the social scientist is to discover and explore those laws (...) Null hypotheses and alternative rival hypotheses developed by social scientists must eclectically correlated to mathematical formulae or the laws of physics in order to advance non-speculative, unbiased knowledge.’
V.J. Belfiglio (p.x)*

So the thesis advanced in *Correlations Between the Physical and Social Sciences* is that social problems can be represented in terms of physical laws. The 41 pages book pushes this argument through four cases studies.

‘The first case study relates marital assimilation of minority groups into dominate core cultures with Graham’s Law for the diffusion of gases. The second case study relates the mutual hostility of political leaders with the Mirror Equation employed in basic geometric optics. The third case study relates the duration of major American military conflicts to the formulae for empirical and subjective probabilities. The fourth case study relates the radioactive decay formula for radioactive substances to the rate of decline of several extinct empires’ V.J. Belfiglio (p.xi)

As the author himself recognizes, “the four case studies in this monograph do not provide definitive answers.” My opinion is on the contrary that they do not provide answers at all. Indeed, the first chapter contains

two 2x2 tables about the endogamous preferences of Mexican and Italian inhabitants of Dallas, Texas (data collected by the author himself). A chi-square test concludes that Mexicans prefer endogamy and that Italians do not. Although Graham's Law is re-expressed there as "marital assimilation being inversely proportional to the square root of the population densities" (p.3), there is no result based on the data supporting this law. The second chapter is trying to "explore the mutuality of hostility between the Bush and Ahmandinejad (sic) administrations. Spearman's Rho correlation coefficient" (p.11) is used and found to demonstrate "a perfect positive correlation" (p.12), although the data is quantitative (intensity of hostility between 1 and 9) and not paired. (The study simply shows that the empirical cdfs of the hostility values for both sides are approximately the same, Spearman's rho test being inappropriate there.) The connection with optics is at best tenuous. Chapter 3 centers on a table for the durations of major American (meaning US) military conflicts. A mere observation is that the US "has been engaged in major wars 56.5 percent of the time between 1775-2010." (p.24) but Belfiglio turns this into "empirical probability" (i.e. the frequency of wars), a "subjective probability" (i.e. the average number of years of peace between wars), and the "number of possible interaction channels" (i.e. a combination number) as a way to link American foreign policy with probability theory. Again, the connection is non-existent. The fourth and final chapter is about the "correlation between the decay of radioactive substances and the rate of decline of empires." (p.31) The data is made of the duration of seven empires, associated with estimates of their half-life. The paper concludes on "a perfect negative correlation between the half-lives of empires and their rates of decline" (p.35), which is not very surprising when considering that one is a monotonic function of the other.

'I conclude with the words of Henry Wadsworth Longfellow: "Sometimes we may learn more from a man's errors, than from his virtues".'
V.J. Belfiglio (p.40)

There is therefore not much to discuss about this book: it does not go beyond stating the obvious, while the connection between the observed social phenomena and generic physical laws remains at the level of a literary ellipse, not of a scientific demonstration. I am deeply puzzled at why a publisher would want to publish it. Any review of the material should have shown the author was out of his depth (his specialty at Texas Woman's University is Government) in this particular endeavor of proving that "mathematical formulae and the law of physics can take scholars further in deriving conclusions from sets of assumptions than can inferential statistics" (back-cover).

Even if the only motivation (of the publisher) for publishing the book is commercial, a mere perusal of its contents is enough to discourage the potential reader...