B Urban Analytics and City Science

Book Review

Environment and Planning B: Urban Analytics and City Science
0(0) 1–2

DOI: 10.1177/0265813516683066
journals.sagepub.com/home/epb
SAGE

Bradley Efron and Trevor Hastie, *Computer age statistical inference: Algorithms, evidence, and data science*, Cambridge University Press: New York, 2016; 475 pp. ISBN: 978-1-107-14989-2, (hbk) $75.00

Reviewed by: Levi John Wolf, Arizona State University, USA.

With *Computer Age Statistical Inference* (CASI), authors Bradley Efron and Trevor Hastie take on two projects. First, they provide a timely, well-written critical history of statistics, discussing how intuition, theory, and computational advances have combined in the development of twentieth-century statistics. Second, the book surveys computational 'data science' methods to show how this development proceeds in the current day. The first two-thirds of the book are exceptional, and the last third is effective but hardly engages with the perspective established by the preceding sections. Regardless, CASI masterfully presents contemporary topics in statistics and (for the most part) embeds them coherently into a valuable perspective on statistical inference as a discipline.

The book is divided into three sections. First, a discussion of 'classical statistical inference' offers an opinionated yet refreshing presentation of the core concepts in modern statistics. The first chapter suggests statisticians frame inference as motivating questions and algorithms as their answers. The rest of CASI presents statistical research by discussing the development of novel algorithms to answer scientific questions, a decidedly 'computer-age' perspective. This places CASI in company with Rey (2009), grounding statistical inference and (spatial) data science on open code, open data, and open inference.

By focusing on algorithms as text, Efron and Hastie exploit the fact that algorithms embed within themselves bits of their author's perspective. Two algorithms that do the same thing may be grounded in entirely different intuitions, and the same algorithm may have multiple justifications. Acknowledging this, the first section closes by establishing three poles of statistical intuition, the Bayesian, the Frequentist, and the Fisherian, each with their own unifying set of arguments and motivations. This tripartite division is a direct reprisal of Efron (1998). Indeed, many of those same arguments are given more room to run across the first two-thirds of CASI. Efron's triangle, a barycentric plot showing where various statistical techniques fall on a Frequentist-Bayesian-Fisherian continuum, later shows up again to visually summarize the map of computer age statistical techniques (p. 265). While a few techniques drift from their spots in Efron (1998), no insight into this movement is offered. The first section's treatment of the fundamentals of statistics is the bare minimum required to move through the remainder of the book, and more thorough training in mathematical statistics would maximize the benefit of the subsequent chapters.

The second section is delightful, illuminating, and helpful. In it, many of the most thoroughly used techniques of contemporary statistics are discussed, like generalized linear models, bootstrapping and jackknifing, model selection criteria, or shrinkage

estimation. Critically, these techniques are presented first as answers to questions *from grounds near their original context*. In some cases, an alternative anachronistic re-imagining from different intuition is provided, but only after the first development. This mode of presentation is a refreshing complement to textbooks where algorithms or estimators are developed 'from nowhere' or more opinionated monographs that re-imagine all of statistics from a given corner of Efron's triangle. This makes the early and middle sections of this book hang well together, united in purpose under a self-aware perspective on contemporary statistical inference. Indeed, the first two sections contain their own epilogue, with the third section calving neatly off the first two with a different purpose, a survey 'more illustrative than definitive' (p. 271).

This 'illustrative' section is packed with frequent references to Hastie's own *Elements of Statistical Learning* (ESL; Hastie et al., 2009). The presentations of cutting edge statistical topics are often more accessible than that found in ESL, and the discussions of large-scale inference and neural networks are particularly useful. But, the first two sections of CASI make a large investment in a coherent perspective about statistical inference that pays less dividend in the third section. The explanations in this section begin to feel like a traditional textbook. By the epilogue, the jump from CASI to ESL feels short indeed. Hopefully, Efron and Hastie update their theoretical perspective in the future to cover this part of their rapidly expanding field. The absence of this extension in CASI belies the uncertainty about statistics as a dynamic and rapidly changing discipline, transforming and enmeshing with opportunities in machine learning and computer science.

Regardless, CASI is useful for its accessible perspective on topics in contemporary statistics. Insofar as it provides context on how inferences are generated from algorithms, the book is a massively useful contribution. In light of this, CASI would serve well in three roles. First, it is a natural backbone for a graduate seminar on topics in contemporary computational statistics, with provided data and code to support instruction. Second, it enriches a PhD statistics curriculum with a balanced point of view on a field that tends to be split between developments from nowhere and strongly opinionated takes. Finally, leaning on detail in chapter end-notes and the source literature, any quantitative researcher interested a deeper understanding of both the theory and intuition of contemporary statistics would benefit greatly from Efron and Hastie's book.

References

Efron B (1998) RA fisher in the 21st century. *Statistical Science* 13: 95–114.

Hastie TJ, Tibshirani RJ and Friedman JH (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. New York: Springer.

Rey SJ (2009) Show me the code: Spatial analysis and open source. *J Geogr Syst* 11: 191–207.