Introduction

Nandita Mitra

Observational Studies, Volume 7, Issue 1, 2021, pp. 1-2 (Article)

Published by University of Pennsylvania Press

DOI: https://doi.org/10.1353/obs.2021.0015

For additional information about this article
https://muse.jhu.edu/article/799740
Introduction

Nandita Mitra  
nanditam@upenn.edu

Department of Biostatistics, Epidemiology, and Informatics  
University of Pennsylvania  
Philadelphia, PA 19104

I am delighted to continue the great tradition that Dylan Small, my predecessor and founding editor of Observational Studies, began of reprinting a seminal paper and inviting leaders in our field to place a lens on it from a current day perspective.

In this special issue we celebrate, examine, and discuss Leo Breiman’s influential paper “Statistical Modeling: The Two Cultures“ (Statistical Science, 2001, 16(3), 199-231) on its 20th anniversary (Breiman, 2001). I remember reading his thought-provoking and somewhat controversial paper when I was a post-doc and it has been a distinct pleasure re-reading it now given the remarkable growth in statistical methodology and specifically machine learning that has taken place over these past 20 years.

We all know of Dr. Breiman’s pioneering work (Classification and Regression Trees, random forests, and bagging just to name a few) that bridged concepts from computer science and statistics and fueled future groundbreaking work in machine learning. His life experiences are presented in Richard Olshen’s fascinating interview in Statistical Science and is worth a read (Olshen, 2001). There I learned that Dr. Breiman not only was a professor at UCLA and UC Berkeley but was also a statistical consultant for 13 years which helped cultivate many of his thoughts on “algorithmic modeling”. Further, he was a sculptor, designed his own home, volunteered in a small village in Mexico, and enjoyed traveling in Africa (Olshen, 2001). Dr. Breiman retired from UC Berkeley in 1993, which was my first year as a graduate student there. I remember running into him in the common lounge space in Evans Hall and feeling too intimidated to speak to him. His somewhat gruff “hello” to me was scary but also something I’ll always remember.

Along with Dr. Breiman’s original paper there appeared four thought-provoking comments by D.R. Cox, Brad Efron, Bruce Hoadley, and Emanuel Parzen. Here, in this special issue, we have 28 new comments written by thought leaders from across the world. I am especially pleased by the diversity of opinions, but more so by the diversity of the authors with respect to gender, race, nationality, professional stage (from graduate student to emeritus professor), job (academia, industry) and field (statistics, biostatistics, computer science, data science, econometrics). Each of these comments stand on their own and will likely fuel more discussion and encourage further research.

I’d like to note that I accepted all comments that were submitted (they were all that good!). In addition, the invitation to write a comment was broadly advertised and open to all. I hope you’ll agree that this helped to provide a broader perspective and allowed for the increased diversity of authors that I mention above. I attempted to categorize the comments into somewhat cohesive topics but decided to publish them in alphabetical order instead (by first author’s last name). However, just to provide a bit of a roadmap
(I apologize to authors who don’t agree with my broad characterizations with the caveat that there is a lot of cross-over of topics), you will find fascinating comments on (1) the blending and cross-fertilization of modeling cultures (not just two distinct ones) under historical, foundational, and flexible paradigms (Olshen; Bickel; Hastie & Tibshirani; Banks; Bühlmann; Kass; Shmueli; McCormick; Mentch & Hooker; Zhao); (2) the importance of interpretable algorithms, understanding the data, outcome reasoning, modeling based on scientific theory, and social responsibility (Eloyan & Rose; D’Amour; Neufeld & Witten; Baiocchi & Rodu; Miller, Foti & Fox; Bradic & Zhu; Freeman et al.); (3) causal modeling (Pearl; Ogburn & Shpitser; Athey & Imbens; Bonvini, Mishler & Kennedy; Cruz-Cortés et al.); (4) Bayesian inference (Gelman; Oganisian & Roy); (5) distributed learning, targeted learning, supervised learning, and computational thinking (He & Wang; Tang, Wang & Zheng; Fan et al; Vansteelandt).

Dr. Breiman passed away in 2005 but I am confident he would have been pleased to know that 20 years later, his paper has inspired so many of us to think deeply about our research, our goals, and our potential for having an impact on the world. I hope you enjoy these commentaries as much as I did.

Settle in with a cup of coffee (or if it’s evening, pour yourself a nice single malt scotch) and enjoy these 28 gems.

Nandita Mitra
Editor, Observational Studies
Professor of Biostatistics
University of Pennsylvania

References
