

Sankhyā : The Indian Journal of Statistics

2006, Volume 68, Part 1, p 167-169

© 2006, Indian Statistical Institute

Book Reviews

Selected Statistical Papers of Sir David Cox (Vol. 1 and 2)

D.J. HAND AND A.M. HERZBERG (EDITORS)

(2005) Cambridge University Press, (xi)+591 pp., (xi)+590 pp.

Price \$ 180.00 (set), ISBN 0-521-84939-x, 0-521-84940-3

These handsomely produced volumes reprint 86 of the more important papers of one of the more important statisticians of the past century. The reproduced papers are presented as originally typeset, and are drawn from David Cox's publications over the period 1950-1993; a full bibliography covering his work from 1947 through 2004 is included at the end of Volume 2.

Collected or selected works of famous statisticians or those who have made important contributions to statistics have a long history. In the 19th and early 20th centuries the works of Gauss, Laplace (twice), Cauchy, Huygens, Chebychev, Fourier, and Lagrange were assembled as reverential monuments, glowing with national pride. In recent times, the works of the ancients, such as Jacob and Daniel Bernoulli and recently Thomas Bayes, have been produced with critical and biographical apparatus as contributions to the history of science. But with modern statisticians, the papers, even older ones, are still a part of the working scientist's reference set, and the motive for collecting them moves towards making them more conveniently available. In some cases (Yule, Neyman, Cochran, Cramér, M.G. Kendall, Robbins, C.R. Rao), they are reproduced with little or no commentary. In others (Wald, Wilks, Wolfowitz, Box, Kiefer, Tukey, T.W. Anderson, Hoeffding), the papers are accompanied by more or less extensive editors' commentaries, to set the place of the papers in the individual's work and the times in which the paper appeared. And in a few cases (R.A. Fisher's 1950 collection and Bartlett's three volumes), the statistician himself has provided individual introductions to each paper, often commenting on the circumstances in which it was prepared. It is the last of these approaches that was taken in preparing the set under review.

The 86 papers presented are grouped in seven thematic areas: Design of Investigations (10 papers), Statistical Methods (25), Applications (6),

Foundations of Statistical Inference (8), Theoretical Statistics (22), Time Series (4), and Stochastic Processes (11). They were drawn from a total of 21 separate journals, 6 edited proceedings volumes, and one encyclopedia; Cox's inaugural lecture at Birkbeck College is also included. The majority of them are from *Biometrika* (23 papers) and three Royal Statistical Society journals (26). Cox's introductory remarks are invariably interesting and perceptive, and they speak well for the editorial decision to adopt this format.

David Cox's extraordinary breadth is strongly represented here, not only the major contributions to statistical methodology (survival analysis, binary regression, data transformation) and foundations (discussions of statistical philosophy, inference, models), but also his modeling of stochastic processes and his insight into what he calls here the design of investigations. The style is distinctive: the ideas are at once clearly expressed and at the same time deep and revealing to even experts. A student could do well to adopt these volumes as a major part of a graduate curriculum! We learn from his introductions about the initial reception of the papers and in some cases how they were developed. Those circumstances do not always provide useful advice to others; for example, "The basic idea of partial likelihood occurred to me while I was feverish." One cost of this approach is that the author's modesty does not convey the full idea of the huge impact some of his work has had. For example, writing of his 1972 JRSSB paper "Regression models and life-tables," Cox tells us it was "cited a fairly large number of times although no doubt read rather less often." This understates the fact that this paper is one of the (if not the) most cited paper in the history of statistics, and in introducing Cox Regression and the proportional hazards model it has had an immensely beneficial effect upon the accuracy and efficiency of a vast array of modern biomedical analyses.

This is not the place or time to attempt to assess Cox's work: even at age 80 he seems still in mid-career, and we may hope that he and the editors and publisher feel obliged to prepare a third volume that includes work since 1993, as well as a few earlier papers that missed out in these volumes, some of the countless penetrating contributions he has made to discussion papers, and an overall index that would enhance the accessibility of the works. When thus completed this will constitute a publishing milestone, a monument to a major body of intellectual work of the first rank and at the same time an invaluable asset for working researchers in statistics. It may also be among the last such sets of printed collected works that will appear, as the costs of production and increasing digital availability make these sets, no matter how useful, uneconomic. Already a bit over three-quarters of these papers are available on JSTOR, for example.

I happen to have recently read through an editorial file for Cox's famous 1958 paper on the foundations of conditional inference, which leads off the second of these volumes. The referees were William Kruskal and Jimmie Savage. In his 8 November 1957 letter to the Editor of the *Annals*, Jimmie wrote: "This paper reflects great knowledge, ability, and experience with controversial questions that are crying out for better understanding today. . . . I predict that it will be read with interest by ninety per cent of the readers of the *Annals*, and I call the other ten per cent intellectual delinquents." The same might be said of these volumes.

STEPHEN M. STIGLER
DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, IL 60637, USA
E-mail: stigler@galton.uchicago.edu