INTERNATIONAL FORUM

DAVID COX: A BRIEF BIOGRAFY[1]

Yadollah Dodge
Institute de Statistique, Universitè de Neuchâtel
Neuchâtel, Switzerland.

David Cox started his career as statistician in 1944 at the age of 20. Here we give a brief story of his scientific life, with extracts from interviews he has given.
David Roxbee Cox was born in Birmingham on July 15, 1924. He studied mathematics at the University f Cambridge and obtained his Ph. D. from the university of Leeds in 1949. He was employed form 1944 to 1946 at the Royal Aircraft Establishment and from 1946 to 1950 at the Wool Industries Research Association in Leeds. From 1956 to 1966 he was reader and then professor of Statistics at Birkbeck College, London. From 1966 to 1988 he was professor o Statistics at Imperial College, London. In 1988 he became Warren of Nuffield College and a member of the Department of Statistics of Oxford University. He formally retires from these positions on August 1, 1994.
In 1947 he married Joyce Drummond. They have a daughter and three sons.
Among his many honours, sir David Cox has received numerous honorary doctorates. He is in particular Doctor Honoris Causa of the University of Neuchatel and is an honorary fellow of Saint John's College, Cambridge, and of the Institute of Actuaries. He has been awarded the Guy medals in Silver (1961) and Gold (1973) of the Royal Society. He was elected Fellow of the Royal Statistical Society of London in 1973, was knighted by Queen Elizabeth II in 1985 and became Honorary Member of the British Academy in 2000. He is a foreign associate of the US National Academy of Science. In 1990 he won the Kettering Prize and Gold Medal for Cancer Research.
At the time of writing this, Sir David Cox has written or co-authored 300 papers and books. From 1996 through 1991 he was the editor of Biometrika. He has supervised, collaborated with, and encouraged many students, postdoctoral fellows, and colleagues. He has served as President of the Bernoulli Society, of the Royal Statistical Society, and of the International Statistical Institute. He is now an Honorary Fellow of the Nuffield College and member of Department of Statistics of the University of Oxford.
He prizes at a personal and scientific level numerous friendships and collaborations built up over the years and specially those with postgraduate students. He taught on the MSc program at the University of Neuchatel for over a decade. Each year the program brought together a mixture of students from many different branches of the applied sciences, and I asked him now how one should teach statistics to a group

[1] Due to the importance of Sir David R. Cox in the contemporary development of Mathematical f Statistics we reproduce in this issue, the article written by Prof. Yadollah Dodge appeared in September 2006 in the Journal Student, Volume 5, number 3-4 pages 153-164. The reproduction was authorized by the editor in Chief of the journal Prof. Y. Dodge.

of not-so-homogeneous students. The following lines are David's random thoughts on the teaching the course on survival analysis, which he was giving at the time:
'' It is always a particular pleasure and challenge to teach away from one' s own institution, the challenge arising in part because uncertainty about the background knowledge that can reasonably be assumed, a key element in the planning of any course. The experience at Neufchatel was notable especially for the impressive variety of subject fields of specialization, experience and interests of the students. This meant that the key issue was not, as so often, that of providing motivation, but rather that of finding a level of treatment that would enable topics close to the forefront of the cotemporary research to the addressed without making unreasonable demands on the students theoretical and mathematical knowledge. Of course, it is hard to judge how successfully this was accomplished.

'There are two extreme strategies that may be used in selecting material for such a course. One is to choose, perhaps rather arbitrarily, two or three topics and to develop these carefully in some depth so that hopefully some understanding of principles is achieved. The other, which I find personally more appealing both when I am a student at other peoples' s lectures, and as a teacher, is to range more broadly and to omit detail.
' This raises the following general issue that arises in teaching any topic that is the focus of much research. Our subject has reached the state where in many fields the literature seems overwhelming. In survival analysis not only are there a considerable number of books, some technically very demanding, but also a very large number of research papers, some wholly theoretical, others concerned at least in part with specific applications, but often also containing new methodological points. Further new paper appear virtually daily
' Yet somehow the explicit message, which I belief to be correct must be conveyed that this volume of work not to be thought overwhelming, that it should be regarded as evidence of the openness of the field and that by examining carefully new applications novel work of value can be done. Put differently there are a limited number o key principles and the rest is detail, important in its proper context, but not usually needed. Thus a central issue in each of the specialized areas of statistics in to separate essential principles from detail
'Much current research in statistics is highly technical; look for example at any issue of the Annals of Statistics. It is sometimes argued: how much easier research must have been all those years ago when simple and important results could be obtained by elementary methods.
' For myself, I very much doubt the correctness of this view, some extent from personal experience, but more importantly the thinking about the history of key developments in our subject.
' Take, for example, the Cramér-Rao inequality. A simple elegant very brief proof is available requiring only the Cauchy-Schwarz inequality, itself provable by totally elementary methods. Does this mean that the pioneers had an easy task? I would be interesting to know Professor Rao's view about this. Certainly the earlier work of Aitken and Silverstone used the mathematically sophisticated ideas of calculus of variations and was much more difficult.
'The tentative conclusion I draw is that statistics was and remains a field in which new contributions can be made for all sorts of backgrounds ranging form those with powerful mathematical expertise to those who have new ideas and formulations arising fairly directly our of specific applications. The intellectual vigour of the subject relies on there being a balanced mixture and thus on our subject attracting people with a wide range of interests. One of the strength of Neuchâtel courses is precisely that.'

' I write some personal thoughts about working as a statistician. I do so largely from UK perspective and from the viewpoint of the statistician, working mostly but not entirely in a university environment and involved with applications in the natural and social sciences and the associated technologies. The nationality aspect is not very important but no doubt had I worked in a government statistical office the issues to be discussed would be very different, although surely the changes there have been great too.

'When I first started statistical work in 1944 there were few people with a specialized academic training in statistics. Most statistical work was done either by those who like me were trained in mathematics, a few of these with some minimal formal training in statistical ideas, or by natural scientists, especially biologists, or by engineers, who had acquired some knowledge of statistics out of necessity. There were very few books, and those that did exist were mostly accounts of analysis of variance, regression and so on in a form for use by non-specialists. They gave little or not theory an hence the intellectual basis for tackling new problem

was often very unclear. In my second job I was constantly faced with sets of data from carefully designed experiments of not quite standard form and I fear I must have done many analysis that were technically based more on hope obtained by extrapolation from simple cases than on understanding.

'Yet paradoxically by that period, i.e. 1945-1950, many of what one now regards as the standard methods of statistics had been developed, but I suppose were not widely known outside specialized groups. Thus Fisher, and very particularly Yates, had developed design of experiments and analysis of variance into highly sophisticated art forms, least squares and virtually all of its properties have even by then a long history. Principal components dates back at least to the early 1900' s and canonical correlation to 1935. Time Series methods, both time and frequency domain had been developed and Neyman had set definite idea n sampling.

' Why were these methods not more widely used? The answer is not so much that they were not known although this may have been a contributing feature, but much more that some were computationally prohibitive except on the very modest scale. A typical statistician of that period would have the help of at least two or three computers. I those days a computer was a person. Such a person could use an electrically operated calculating machine to do standard balanced analysis of variance, simple regression calculations with up to two or three explanatory variables, simple chi-squared tests and maybe some other specialized calculations. There was also a rich variety of special types of graph paper for various graphical procedures. Of course other calculations could be done but would need special planning.

'The possibility of numerical errors was ever present. Important calculations had to be checked preferable by a distinct route. I had the enormous good fortune to work for a period with Henry Daniel. He was also meticulous. Numerical wok had to be done with a sharp pencil entering the answers two (never three) digits per square on the quadrille paper (ruled I ¼ inch squares). Errors had to be rubbed out (never crossed out) and corrected. Tables etc. had to have verbal headings.

'Probably the most distressing calculation was for time series analysis. The data were rounded to two digits an entered twice into a comptometer, an old=style cash register. From this one printed out a paper strip containing the data in columns, with each value occurring twice. This was cut down the middle to produce two strips. Paper clips were then used to attach these together at displacements h, the desired log for a serial correlation. Corresponding values were subtracted mentally and then squared mentally and the answer entered usually by a second person, into an adding machine. This produced …(yi+h-yi). From which the lag h serial correlation can be obtained. This would be repeated for as many values of h as were required, constrained by the need to preserve the sanity of the investigators. Of course I take for granted knowledge of the squares of integers up to say 50 or so.

'Matrix inversion takes a time roughly proportional to the cube of the size. The largest I ever myself was, I think, 7x7which took; with obligatory checks most of one day.
' Now the consequence of all this for the analysis of data was: look very carefully to the data, to asses its broad features, possible defects and so on. Then think. Then compute.

' The idea of an electronic computer was known ad I recall in particular the visionary statements of I. J. Good and A. D. Booth, who foresaw the impact computers would have in daily life. It is interesting that it is taken almost 50years for what they anticipated to come to pass. I suspect that one if not both of them assumed gentler future in which the miracles made possible by scientific progress would enhance the general quality of life, not to be used as the driving force behind an ever more frantic and in many ways competitive scramble.

'In statistics, as a sweeping oversimplification, the period up to 1975-1980 was taken with getting the classical procedures, and their direct developments, into a form which they could be relatively painlessly used. Since then the new horizons opened by massive computer power, including the ability to handle large amounts of data, are being explored.
' The consequence of all this for the analysis of relatively modest amounts of data by contrast with the attitude summarized earlier, is: first compute. Then think.

‘In one sense this is a liberation and I do not doubt that the balance it I good thing, but there is a downside on which we all need to reflect.

‘ How different now is the statistical wok form how it was then I started. First there are, of course many more individuals with statistical training and much deeper percolation of statistical ideas into subject matter fields. Think for example of medical research. I many, although certainly not all areas the specialized subject-matter medical research journals hinge on the use of quite advanced statistical procedures in a way that was unthinkable even 20 years ago.
‘ The scope for discussion with and help and advise from knowledgeable colleagues is far greater than it was; few statisticians work totally on his own. I remarked above on the virtual no existence of books when I started. Now the issue is more the most overwhelming extent of the literature and not least the appearance each year of so many books, the great majority of a very good standard. Of course the technical advance of the subjects leads to statistician becoming more and more specialized into particular areas.
‘Yet in some sense I wonder of things have changed all that much a deeper level. Really serious subject-matter research problems virtually always have unique features with a demand to be talked, to some extent at least, from the first principles and for this sound knowledge f the central ideas of subject and an appropriate attitude to scientific research requires more than very specialist information. My own work has been held back more often by insufficient subject matter and, especially in my earlier years, by an inadequate theoretical base in statistics and insufficient mathematical depth, rather than by unfamiliarity with particular specialize techniques.

‘Although mostly research nowadays is done in groups rather than by isolated individuals, research, whether theoretical or applied, remains an intensely personal activity and the best mode of working has to be found by trail and error by each one of us. The subject is at an exciting and demanding phase and I envy those starting on a career in statistical work.’

And finally, here is the humble of a real scientist, provoked by two questions asked by Nancy Reid in her conversation with David Cox.

‘The only other thing I wanted to ask you about is something that you mentioned to me in a letter a while ago which I think was connected with your knighthood. Your words were roughly that, after feeling when you were younger that you didn’t get very much recognition for your work, you now felt were receiving a ‘’bizarre excess’’,
David Cox: ‘Yes, I think that sums it up adequately. Well, everybody needs encouragement, and of course as you get older you still need encouragement. But the time you most need it is when you are starting. It would be quite wrong to think that people were ever discouraging, they weren’t. It was all very low keys, typically British understatement sort of thing. You never really knew what people thought, despite the relative frankness of Royal Statistical Society discussions. And I’d published a few papers with a very little notion of whether anybody had paid attention of them until I first went to the United States, where people would come and say, “Oh, I read that paper, I think we could do so ad so.” That sort of thing is very encouraging. And then, more recently, I’ve been absurdly lucky with all thee pieces of recognition it’s more likely you’ll get another. It ought to be way around.’

‘Was there a time when you suddenly felt you were a Very Important Pearson?’
David Cox: “Well, I hope I’ve were thought in those terms. In a sense, the only thing that matters is if you can look back when you reach a vast, vast age ad say, “Have I done something reasonably in accord with my capability?|" If you can say yes, okay. My feeling is in one sense, I’ve done that: in the tangible sense of books and papers, I’ve done more than I would have expected. In another sense I feel very dissatisfied: there are all sorts of problems that I nearly solved and gave up, or errors of judgments in doing a little something and not taking I far enough. That I nearly did something you see, this is the irritating thing. You know, if you’d no idea at all, will it doesn’t matter, it’s irrelevant, but if you feel you were within an inch of doing something and didn’t quite do it…