Res. Lett. Inf. Math. Sci., 2005, Vol.**8**, pp45-52
Available online at http://iims.massey.ac.nz/research/letters/

Recollections from a 50-year random walk midst matrices, statistics and computing

Shayle R. Searle

Department of Statistical Science
Cornell University
Ithaca, New York, USA

A brief and personal overview is given of developments in matrix algebra, statistics and computing during the years of my participating in these activities, 1945 2005.

1 Education

First a word or two about my education. Even in my early schooling I always liked mathematics; so in 1945 I aspired to becoming an accountant. But six months as junior office boy in a large accountancy firm in Wellington, where I was introduced to auditing (checking the page by page totals of a company's books) soon cooled those aspirations. So in 1946, at Victoria University College (one of the six then colleges – now all universities – of the University of New Zealand) in Wellington I started a degree in mathematics. My aim was to become an actuary. In those days mathematics graduates had limited choices for a career in New Zealand: school teacher, meteorologist, accountant and actuary. Only the last of this list attracted me and seemed possible. So after the M.A. degree I became an actuary's assistant in a Wellington life insurance office. Then, in 1951 I went to England (31 days by passenger ship – deadly dull except for a hurricane in the Caribbean) ostensibly to do statistics at Cambridge and then to pursue actuarial qualifications.

But the years 1950 – 51 had some knock-backs. In approaching the M.A. exams I wanted to apply for an overseas scholarship: this had to be done before the exams. However, I was not allowed to apply, because my professor had to sign the application and would not do so: he told me I was not good enough. (Imagine that happening in today's world!). Nevertheless my exam results confounded the professor and I could have applied the next year, but by then Peter Whittle was finishing his Master's degree – so for me that was nolo contendere. Despite all this I decided to go to England anyway. In the meantime I prepared for the Part I exam of the (London) Institute of Actuaries, and this was knock number two: two hours before boarding ship for Southampton a phone call informed me that I'd failed the whole exam.

Following a few days in London I went up to Cambridge and on getting to the Statistical Lab, its director, John Wishart (of Wishart distribution eminence) greeted me with "who are you? I've never heard of you!" The admission's bureaucracy had gone astray! But Wishart had a most gracious Englishman's reaction: "Well", he said, "you've come a long way, we can't send you back!" So I stayed.

For academic potentials (which I did not fulfill) I was allocated to Dennis Lindley as an adviser. He soon saw my deficiencies and after three months advised me not to pursue a PhD. Reason: In 1951 the gaining of a PhD depended on examiners from the mathematics department and if they were not sufficiently impressed at the oral exam (no course exams) the outcome was a Master's degree – which, under those circumstances meant "failed PhD." So to avoid this unwanted label I took Lindley's advice and worked for the Diploma in Mathematical Statistics. It provided course work, and being farmed out to a university department that had data awaiting analysis. I was sent to Dr E.H. Callow whose data were the percentages of bone, fat, muscle and lean meat in beef carcasses. One of the three final exams for the diploma was data analysis, using hand-operated mechanical (not electric) Brunsviga calculators. Their random clattering cacophony of screeching wheels did not make the exam any easier.

Well, I passed, but only just and certainly without the "with distinction" that Lindley was expecting of me. Wow! Did I get a dressing down! However, I'd had a good time over the two years.

From all of this I have a philosophy: never let a knock-back be a knock out. Don't give in. In today's world I find a tendency toward making education easier and to wanting people not having to face up to the tough moments in life. I think that's wrong. We need to learn to handle the difficult moments in our individual lives. Each of us has them – right up to old age, disabilities and death of loved ones. Students today want no examination pass list publicized. "I don't want my friends to know I've failed." Why not? Face up to it, find an alternative and get on with life.

Statistics was a relatively new academic discipline in 1951. There were few books: 18 inches of shelf space in Lewis bookstore near London University included M.G. Kendal's first volume (some of it written in a wartime air raid shelter), Cramer, Fisher and Good. Lectures at Cambridge were from H.E. Daniels on applied statistics, Wishart on combinatorics (fun but of little practical value, D.R. Cox on design (using Kempthorne's 1952 book), probability from Lindley and something from Anscombe. What a star-studded group of young lecturers – and Cambridge let them all go (save Wishart, who was a Reader in Agriculture, and who drowned in the ocean off Mexico).

Exams done, the Diploma led to job hunting. The career center at Cambridge was not accustomed to post-graduate statisticians, least of all New Zealand ones! Nevertheless I did survive two very formal interviews and got offered jobs: one was with the British Colonial Service in Nairobi where I would have had my own jeep, driver, Marchant calculator and rifle; the other was to be with Royal Dutch Shell in Caracas, Venezuela. I was too timid to take either.

2 Dairy Cows

So I returned to New Zealand, missed out in my application for a position at Ruakura Research Station, and landed up in 1953 as Research Statistician with the New Zealand Dairy Board in Wellington. The work was with production (milk weight and percent butterfat) records of the many thousands of dairy cows whose such records were being collected monthly in what is known as a Herd Improvement Program. These records were (and still are) wonderful data for statistical and genetic studies as, for example, in selecting bulls for use in artificial breeding and for estimating heritability (a widely used genetic parameter). It is the ratio of genetic variance to phenotypic variance. This demands estimating variance components and doing so from data having unequal numbers of observations in their subclasses, now called unbalanced data. This was dealt with by Henderson (1953) in three variations of what nowadays is known as the analysis of variance method (of equating sums of squares and other quadratic forms to their expectations). But Henderson's paper had a big hole in it: nothing about the sampling variances of the variance components estimators. That really set me going: what are those sampling variances? They are second degree functions of the true variance components, having coefficients that are horribly complicated third and fourth degree functions of the (unequal) numbers of observations: Details are in Searle (1956 and 1958). (A colleague's later comment was that they are probably the only papers in the Annals of Mathematical Statistics that have an author's address of dairy farming origin!). Unfortunately, the expressions for the sampling variances were not only difficult to calculate but were totally impractical for comparing their behaviour for different data sets.

September 1955 was a hallmark month: C.R. Henderson, Professor of Animal Breeding at Cornell came to New Zealand for a sabbatical year. He was allocated a desk in my office (which had been the large bedroom of a big old house in Bolton Street, Wellington).

Henderson was finalizing his development of BLUP (best linear unbiased predictor) – and with the assistance of some matrix algebra (inverting a partitioned matrix) confirmation of his conjecture was firmly established. Fired by this result, and having him as office mate, I soon got to know him and to realize that his prime interest in animal breeding was statistical. So I declined acceptance to the PhD program at North Carolina State University and in 1956 went on a Fulbright travel grant to Cornell to work under Henderson. Course work in breeding consisted mostly of dealing with records of related animals, and in statistics the mainstay was analysis of variance.

3 Statistics

Expressing model equations and sums of squares in matrix terms had barely begun: Kempthorne (1952) had made a start but it bred little follow-up. What was so lacking was knowledge of estimability which Bose was developing in North Carolina in the 1950s. In 1957-8 Henderson and I presented a weekly seminar on analysis of variance, normal equations and estimation of effects in linear models. Throughout it all, and the 200 pages of notes which it produced, there was always confusion about estimation in the presence of constraints – such as those where effects are defined as adding to zero (or one as being zero). The consequences became all so clear after Rao (1962) had brought to statistics the generalized inverse matrix of Penrose (1955). Rao clearly showed how it can lead to numerous different solutions of a set of equations not of full rank such as so often occur with normal equations in the linear model. And this leads to certain linear combinations of the elements of any solution vector being invariant to the solution vector – and to being the unique BLUE (best linear unbiased estimator) of a certain linear combination of the effects in the model.

4 Computing

Then, in the mid 1950's, along came the start of stored program computers. Cornell got its first in 1956, enthusiastically promoted by Henderson. He could see computers would lead to farmers receiving valuable monthly production data on each of their cows. How right he was! That first computer at Cornell was an IBM 650, having a rotating drum with 2000 addresses for storing 10-digit numbers (with sign). Initial programming was with BLISS (Bell Labs Interpretive System), but that was soon supplanted by SOAP (Symbolic Optimum Assembly Program). Taking the drum's rotation into account SOAP put each instruction on the drum in an address optimal (in terms of speed) to the address of the preceding instruction. And, of course, as the drum got to contain more and more instructions the optimality declined. But it was much more efficient than doing the optimizing oneself. And the outcome seemed to be amazing. Inverting a 10 by 10 matrix in 7 minutes left us speechless – especially compared to inverting a 40 by 40 in 6 weeks, as occurred here in New Zealand in the pre-computer days.

With my PhD completed I had a third year at Cornell, in Henderson's group, analyzing dairy records from New Zealand. On a 1959 visit to IBM in Endicott, N.Y., we heard about FORTRAN: it consisted of 6,000 cards and involved a 3-pass procedure, via FORTRANSIT to get a FORTRAN program as output. Those 6000 cards came with me on returning in 1959 to my position with the Dairy Board in

New Zealand – but we never used them!

In 1962 I was invited back to Cornell as statistician in the university's computing center which was on the point of getting a CDC 1604 computer, which came with printer attached! Big deal! There was as yet no commercial statistical software available; just starting was BMD out of UCLA, under the direction of Will Dixon, (of the Dixon and Massey text) and supported by funds from N.I.H. SAS had, I believe, not even started (it's first user's conference was 1975). It began at North Carolina State University with graduate student Jim Goodnight developing an analysis of variance program for Southern Experiment Station Statisticians.

So, without software, my responsibilities were two fold: to offer statistical consulting to students and faculty, and to plan what programs we should write (I had a programmer available) to bolster this consulting. We concentrated on analysis of variance, multiple regression, contingency tables and rank correlation. An early consulting job illustrates how primitive (compared to today) our usage of the computer was. The author of a published analysis of variance was very concerned at receiving correspondence from colleagues suggesting that his analysis must surely be wrong. What could he do? I asked to see the data: some five hundred of them. The client asked "What will you do?". "Wait and see" I said as I took from my pocket a coin to use as a place-holder while starting to slowly slide it down the column of data. "How will that help?" I was asked. "Let's see" I replied, and sure enough we soon come across two data values that had been mis-punched on the IBM cards of those days. They had been punched two places to the left of where they should be and so were 100 times too big. Problem solved. Using the computer to edit data had scarcely begun. And when it did, an early procedure was to record a missing value as -1. But at least one student then did his analysis of variance treating all the -1's as real data!

5 Linear models

In 1965 I moved from Cornell's computing center to its Biometrics Unit on a regular professorial track. Computing politics were in the throws of centralization verses decentralization, and I wanted out. I sat in on a poorly organized course from Graybill's (1961) book on linear models: in six weeks it had got no further than quadratic forms. Two years later a colleague offered the course but after a month dumped it on my lap! I was appalled by Graybill's treatment of quadratic forms: a succession of some twenty theorems, each having a minor change (and proof) dependent on its predecessor. The very last was an important and all-embracing theorem: conditions for a quadratic form of normally distributed variables to have a non-central chi-squared distribution. It seemed to me to be a much easier teaching and learning plan to deal with that important theorem first as a strong, broadly applicable stand-alone entity and then show that many of the Graybill theorems were just special cases.

So all this got me started – first, on a single sheet of paper, by summarizing each theorem on a single line. This made it easy to see the differences between the theorems. No doubt those theorems were useful for students who would be asked to regurgitate some of them in closed-book exams. However, that was not my desire in teaching. I wanted students to get well equipped in understanding the theory, to know where the technical details were written down, and to be able to thoroughly understand them whenever reading them. After all, in occasionally applying theorems to practical situations one cannot always risk relying perfectly on one's memory, but one must know (1) where the details are written and (2) how to read, understand and use them.

This outlook was the beginning of my 1971 book *Linear Models*. A popular, but by no means universal, concept of analysis of variance at that time was that of simply partitioning a sum of squares into a sum of two or more other sums of squares, as Fisher had promoted. But Urquhart et al (1973) showed that Fisher had used no linear model equations. Hazel (1946) had analyzed lamb weights "defined by (an) equation" but had not called it a model equation – nor used it to calculate any sums of squares due to sub-models. Various authors had used model equations as under-pinnings for finding expected values of sums of squares for estimating variance components, and Henderson (1953) had dealt with this for unbalanced data. But not much had been done for analysis of variance for unbalanced data based on fixed effects models. Yates (1934) had described the weighted squares of means analysis, but it was restricted to data with no empty cells, all-cells-filled data as they are now called. Furthermore, there was no involvement of the concept of estimability, namely which linear combinations of effects could be estimated and how to estimate them. And the handling of interactions was not clear – nor was there a uniform procedure for doing so, especially for data with three or more factors.

Then along came SAS (Statistical Analysis System with its GLM (General Linear Models) software. In 1975 at the first SAS user's meeting their four types of sums of squares were described. At an informal after-dinner session a bunch of we linear model gurus were asked which of the four types should be the printed output. No unequivocal answer was forthcoming so, understandably, SAS stayed with all four.

6 Matrices

The big impetus to the development of many results in linear models, multivariate statistics and other specialties is, of course, matrices and their algebra. The reason is that any data list can be represented as a vector and although the traditional partitioning of sums of squares was suited to data from well designed and executed experiments, giving balanced data, or planned unbalanced data such as from latin squares, the advent of computers motivated having algebra that was free of the amount of data. Matrix algebra achieves this.

Nevertheless, although matrices were available before 1850, their full infiltration into statistics has been very slow. Searle (2000) provides extensive evidence of this. Just a few examples: Wicksell (1930), the first paper in the first issue of the

Annals of Mathematical Statistics, is titled "Remarks on Regression" – yet it has no matrices. Neither did the teaching of multiple regression in Cambridge when I was there in 1952! And in Aitken's two 1939 books, *Determinants and Matrices*, and *Statistical Mathematics*, neither mentions the topic of the other, save for a snippet about quadratic forms in the matrix book. Even later, the book of Williams (1959) on regression has only a tiny mention of matrices. And in only its most recent edition has Snedecor's widely used book acknowledged the use of matrices. But today they are an essential tool for describing and developing many statistical procedures and for designing the welter of software that has become available for the otherwise nasty arithmetic.

7 Teaching enough mathematics?

But, and this brings me to a final point. The decades from 1930 until 1990 were a time of enormous development of statistical methods, achieved through ingenious algebra done by people who presumably enjoyed doing it – they enjoyed mathematics. Today the tedious arithmetic which is the result of that algebra, and even some of the algebra itself, can be done by software. But in teaching how to use that software and it's output, I fear that we may not be teaching enough of the underlying algebra; and that we are trusting the software infallibly. Certainly most established software is correct, most of the time, maybe as much as 99.7% of the time. But in the remaining 0.3%, or even 0.03%, when a software output is a nonsense value who will have the training to trace its origin and correct it? To do that requires a solid knowledge of the appropriate mathematics. Even more seriously who will have that solid knowledge for developing new methods?

But is our teaching doing a good job of identifying students who like and enjoy mathematics? They are likely to be the people who find it fun to do mathematics, and so go on to make good use of it. For myself, learning that with matrices AB could be null without either A or B being null absolutely fascinated me – and I became hooked on matrices.

So now what of the future for matrices, statistics and computing? Matrix studies seem to become increasingly difficult, statistics has enormous opportunities for careers, but needs more and better public relations; and computing is racing ahead so rapidly that one has difficulty trying to keep up with it. Other than saying that, I'm not willing to even try to predict the future: we all know that prediction outside the range of our data is a risky game.

References

Hazel, L.N. (1946) The covariance analysis of multiple classification tables with unequal subclass numbers. Biometrics Bulletin 2, 21-25.

Henderson, C.R. (1953) Estimation of variance and covariance components. Biometrics 9, 226-252.

Kempthorne, D. (1952) The Design and Analysis of Experiments, Wiley, New York.

Penrose, R.A. (1955) A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51, 406-413.

Rao, C.R. (1962) A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. Journal of the Royal Statistical Society, Series B, 24, 152-158.

Searle, S.R. (1956) Matrix methods in components of variance and covariance analysis. Annals of Mathematical Statistics, 27, 737-748.

Searle, S.R. (1958) Sampling variances of estimates of variance components. Annals of Mathematical Statistics, 29, 167-178.

Searle, S.R. (1971) Linear Models Wiley, New York.

Searle, S.R. (2000) The Infusion of Matrices Into Statistics, Image 24, 25-32.

Urquhart, N.S., Weeks. D.L., and Henderson, C.R. (1973) Estimation Associated with Linear Models: A Revisitation. Communications in Statistics A: Theory and Methods 1, 303-330.

Williams, E.J. (1959) Regression Analysis, Wiley, New York.

Yates, F. (1934) The analysis of multiple classifications with unequal numbers of observations in the subclasses. Journal of the American Statistical Association 29, 51-66.