Statistics and the Computer

Lennart Råde - Gothenburg, Sweden

1. Introduction

It is common to describe the impact of the computer on statistics and on the teaching of statistics as a revolution.

> "The advent of the computer has certainly revolutionised the practice of statistics." (Speed, 1985).

> "It is clear that both the theory and practice of statistics are being revolutionised by the computer and that, as a result, radical changes are taking place in the teaching of statistics." (Lunn, 1985)

Sometimes these views are modified by referring to several successive revolutions or by reference to different kinds of computers. (See, for example, *Computers, the second revolution in statistics* (Yates, 1966) and *Microcomputers : The coming revolution in statistics* (James, 1983).)

The recent conference SOFTSTAT '89, the 5th conference on the scientific use of statistics software, held in Heidelberg, Germany in April 1989, is a concrete indication of the interest in and the importance of computers in statistics. The conference was attended by about 500 participants and the Proceedings (Faulbaum et al., 1990) contain more than 70 papers totalling over 600 pages. The following is a list of statistical software packages which are mentioned in these Proceedings.

AGREE	BLOGIT	BMDP	CensSys
CHOMP/CONFID	CHISQ	CLINTRIAL	CONF
DATADESK	DISTAN	EGRET	ELASTIC
EQS	EXPLO	FASTAT	Gauss
GENSTAT	GLAMOUR	GLIM	GSTAT
HLOGIT	IMPOS	JMP	LACORD
LIMDEP	Lincs	LISCOMP	LISREL
MANYEARS	MAPLIS	MDA	MicroTSP

MIMOSE	PC-ISP	PES	P-RADE
P-STAT	PYTAB/AMFIT	S	SAS
SCATTER	S-PLUS	SPSS	STATGRAPHICS
Statistik-Prakt	Statlab	STATPAL	STATPRAK
Statview	STEM	Survo	SUSA
SYSTAT			

2. Statistics and the computer

There are two different ways in which the computer is changing the field of statistics. First, computers can help us to do what we did before the advent of the computer but in a more efficient way. Second, computers can help us to do things nobody thought of before the advent of the computer.

To the first category belong statistical data analysis by numerical and graphical methods, and simulation; to the second belongs, for example, different computer-intensive methods (see Diaconis and Efron, 1983). Another way to categorise the relation statistics-computer is to list the different ways the computer can be used in statistics. The following are examples of such uses: numerical and graphical data analysis; symbolic computations; simulations; storing statistical knowledge; presentation of results.

The close relationship between statistics and computing implies that when one changes the other will also change. The following are some new practical procedures in computing which have turned out to have a great importance for statistics:

(i) The change from mainframe batch computing to personal computing.
(ii) The introduction of multiple dynamic displays.
(iii) The possibility of direct manipulation of graphical objects.

Some trends in statistics are also obviously very much influenced by what has happened in computing. Examples of such trends are:

(i) emphasis on exploratory data analysis instead of hypothesis testing;
(ii) the use of computer-intensive methods;
(iii) the introduction of new diagnostic methods.

3. Teaching of statistics and the computer

We are at a conference on the teaching of statistics. It is therefore natural to try to discuss in more detail the opportunities offered by the computer to improve the teaching of statistics.

It is, of course, necessary to consider the use of computers in every course in probability and statistics, both at school and university levels. However, the introduction of computers in the teaching process has, in many places, been rather slow. One obvious reason for this is the considerable cost of establishing an effective computer environment to use in the teaching process. Other reasons for the slow

development are the many difficult problems to be solved before the computer can be incorporated in the teaching of statistics. Some of these are:

(i) What kind of computer should be used (mainframe, personal, pocket)?
(ii) What make of computers should be used?
(iii) What software should be used?
(iv) What programming language should be used?
(v) What teacher education is needed?

Here only points (iii) and (v) in the above list will be discussed.

There exists a tremendous amount of statistical software (see the list in the introduction above). These software packages are usually produced for use in statistical practice rather than for teaching. But very often they can be used for that purpose, especially if they fulfill the following criteria:

(i) are easy to use;
(ii) have good supporting material;
(iii) are menu-driven;
(iv) have colour;
(v) have good graphics;
(vi) are interactive;
(vii) have relevant content.

Courses at university level should give the students familiarity with some common statistical packages, including the new kind with dynamic graphical facilities (e.g. DATADESK, JMP). One possibility is to have a specific course on statistical computing which includes the use of statistical software, but it might be better to incorporate the use of the computer in all courses in probability and statistics. It can be mentioned here that at some universities in the United States, complete undergraduate programmes in computational statistics have been introduced.

There exist very few educational programs in statistics. The making of such programs is discussed at length in Schleiffer (1989), where such programs are classified into three major groups: programs for drill and practice, programs for simulation, and programs for modelling. Schleiffer has developed three educational programs (Confidence intervals, Coin tossing, and Tests) to be used in the special computing environment (one Macintosh computer available for each five students at the university) at ETH (Eidgenössische Technische Hochschule) in Zurich, Switzerland. Among other educational programs in statistics can be mentioned Statistik-Praktikum (Afflerbach, 1990) and GSTAT (Böker, 1990). Computer illustrated texts for teaching statistics are discussed in Bowman and Robinson (1990).

Another problem is that we lack a suitable system for assessing and controlling the quality of statistical software, especially how useful statistical software is for teaching purposes. This problem is discussed at length in Molenaar (1990).

Teacher training in the use of computers in teaching probability and statistics is a serious problem, both at school and university levels. Even teachers with a recent degree in statistics might not have the proper background in computational statistics and it is in any case not clear how to incorporate the computer in stochastic courses. It is

thus important that institutions are given enough funds to support teacher training in this area. It is also to be noted that the introduction of computers in the teaching of stochastic courses is a very time-consuming affair.

The case of ETH in Zurich, Switzerland, can serve as a good example of what is required at the university level to promote and facilitate the use of computers. When ETH established its computer facilities for the students, it also established an institute IDA (Informatik Dient Allen) with a staff of seven with the task of helping in the process of introducing computers into university courses.

4. Software for symbolic calculations

A new important feature of computation is the advent of software for symbolic calculations. These computing packages (e.g. Macsyma, Reduce, SMP, Derive, Maple, Mathematica) are intended for mathematical calculations but they can turn out to be extremely useful in probability and statistics as well. A package like Mathematica offers a large number of mathematical tools for both numerical and graphical computations and contains a lot of useful mathematical knowledge. It also contains some statistical packages for statistical data analysis, with information on statistical distributions. The number of such packages will probably increase rapidly in the future. These packages can perform algebraic manipulations, integration, differentiation, finding and inverting Laplace and Fourier transforms, linear algebra operations and so on, all of which are important in applied probability and statistics (some examples will be presented in the lecture at ICOTS 3).

5. The future

This is an exciting time for statistics and for the teaching of statistics. Modern technology offers new very useful tools both for the theory, the practice and the teaching of our subject. Extremely powerful software for probability and statistics theory and practice and also software for making software will appear in the future and will be readily available on personal computers. (Expert systems have not been discussed in this paper as the author lacks experience of such systems.) Parallel computers will be of importance, for example, for simulations.

The computer will surely find its way into the classroom. As we today take it for granted that a classroom is equipped with an overhead projector, so will it in the future be taken for granted that classrooms have access to computers both for student use and for demonstration of software action. Some teachers might prefer to do all their teaching with the computer, but there will probably be some place for chalk and blackboard as well. The computer will be the daily working environment both in our theoretical and practical work.

In the following statement from the report by Demana et al. (1990), the word "mathematics" has been replaced by "statistics".

> "In the final decade of this century, paper and pencil will take its last stand in statistics."

References

The following list contains references to works mentioned above and also to some other reports, books and papers. A special reference should be given to the report by Biehler (1988) which contains a more complete list of references to works in this field. See also Schleiffer (1989).

Afflerbach, L (1990) Statistik-Praktikum mit dem IBM PC. In: F Faulbaum, R Haux and K H Jöckel (eds) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

Biehler, R (1988) *Computers in Probability Education.* Occasional Paper 108, Institut fur Didaktik der Mathematk der Universitat Bielefeld.

Biehler, R and Rach, W (1990) Software tools for statistical data analysis in education and teacher training? In: F Faulbaum, R Haux and K H Jöckel (eds) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

Bowman, A W and Robinson, D R (1990) Teaching statistics : Microcomputer graphics and computer illustrated texts. In: F Faulbaum, R Haux, K H Jöckel (eds) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

Böker, F (1990) On the statistical package GSTAT. In: F Faulbaum, R Haux, K H Jöckel (eds) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

Demana, F, Watts, B K and Harvey J (1990) The twilight of the pencil and paper. *Proceedings of the Conference on Technology in Collegiate Mathematics.* Addison Wesley.

Diaconis, P and Efron, B (1983) Computer-intensive methods in statistics. *Scientific American* **248**.

Faulbaum, F, Haux, R and Jöckel, H (eds) (1990) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

James, E B (1983) *Microcomputers : The Coming Revolution in Statistics.* Invited paper at the ISI Session in Madrid.

Lunn, D (1985) Computer animation : a powerful way of teaching concepts of probability and statistics. In: L Råde and T P Speed, *Teaching of Statistics in the Computer Age*. Chartwell Bratt Ltd., London.

Molenaar, I W (1990) Producing, purchasing and evaluating statistical scientific software : intellectual and commercial challenges. In: F Faulbaum, R Haux, K H Jöckel (eds) *SOFTSTAT '89, Fortschritte der Statistik-Software 2*. Gustav Fischer Verlag, Stuttgart, New York.

Schleiffer, C (1989) *Some Educational Programs in Statistics.* Research Report No 61, Seminar fur Statistik, Eidgenössische Technische Hochschule, Zurich.

Speed, T P (1985) Teaching statistics at the university level : how computers can help us find realistic models for real data and reasonably assess their reliability. In: L Råde and T P Speed, *Teaching of Statistics in the Computer Age*. Chartwell Bratt Ltd., London.

Yates, F (1966) Computers, the second revolution in statistics. *Biometrics* **22**.