

Computational Statistics or Statistical Computing, is that the question?*

Carlo Lauro

Dip. Matematica e Statistica, Università di Napoli, Via Cintia, I-80126 Napoli, Italy

Past President of IASC

During last COMPSTAT Symposia, a recurrent matter of debate regards the meaning, the contents and the future of this recognized and successful meeting that has come by now, to its XII edition.

A crucial point of this debate concerns the definitions of Computational statistics and statistical Computing. Actually, the lack of an agreement on their meanings allows for confusion in the debates on the contents of both COMPSTAT and the IASC journals.

The two terms are often used as being interchangeable or mutually inclusive. Sometimes, they are understood as being different, with the former meaning a more theoretical or academic approach of the European School, with respect to the more applicative one of the latter, followed mainly by the North American School.

From an historical point of view, the birth of the conference activity in the area of interest may be dated in 1967 when the Interface Symposium of Computer Science and Statistics took place in Atlanta. The term Statistical Computing has been later adopted by the two major scientific associations in this area, i.e. the Statistical Computing section of the American Statistical Association founded in 1972, and the corresponding section of ISI known as International Association for Statisti-

cal Computing. The latter, founded in 1977, included the previous European group known as COMPSTAT, which started this series of biannual scientific meetings with the first one being held in Vienna in 1974.

The Statute of the section of Statistical Computing of ASA reports as objectives "Encouraging the application of computer hardware, software and systems to statistical problems, ... Encouraging research in statistical computing and communication of the results". The objectives of IASC are "to promote the theory, methods, and practice of statistical computing and to foster interest and knowledge in effective and efficient statistical computing through international contacts."

As it can be seen, both associations rely to the application of computers to statistics but IASC, differently from ASA, puts a major accent on the development of the theoretical and methodological aspects of Statistical computing. Does this point suffice to justify the names COMPSTAT and CSDA (official journal of IASC) as an expressed preference of IASC towards more theoretical aspects (Computational Statistics)? Such a feeling would not be supported by the term recently adopted by the Eastern Asian Regional Section of the IASC (EARS) which refers to its conference,

corresponding to the COMPSTAT of EARS, as STATCOM.

Therefore, do we have to believe that Computational Statistics and Statistical Computing are considered as being interchangeable by IASC? It would be so if we looked just at the definition given by Norbert Victor, who states the "Computational statistics is the examination of possible solutions and simplification of the statistician's problem by the application of computer and/or the method of Computer Science" (Computational Statistics - Tool or Science? SSN, 3, 1984, 105-116).

This view is no more supported when looking at the most accepted general purpose journals in the area, Computational Statistics and Data Analysis (S. Azen), Computational Statistics (W. Härdle), and Statistics and Computing (D.J. Hand). Their respective aims and scopes, but for the specific reference to Data Analysis of CSDA, shows a list of very similar topics ranging from the design of algorithms to implement technical methods, to the development and evaluation of statistical systems, so the newest hardware and software environments in Statistics, as well as the application in Statistics of the most recent technologies of Computer Science.

It is easy to understand that the above topics have an impact on the different areas of Statistics. They all contribute to define, besides any opposition between Computational Statistics and statistical Computing what I like to call "Statistics in Computer Age" (see my discussion to the paper of N. Victor in SSN, 3, 1984, 119-120).

Nevertheless, the titles of the very few reference books in the area are helpful in clarifying the debate. Kennedy and Gentle (Marcel Dekker, Wiley, 1980), in fact, prefer the term Statistical Computing, Thisted (Chapman and Hall, 1988) adopts the term "Computational Statistics", whereas Maindonald (J. Wiley, 1984) is for "Statistical Computation". Although he different titles, the books are characterized by a strong common structure mainly based on numerical methods and the algorithms for solving statistical problems (random numbers, computational methods in linear algebra, optimization methods, smoothing and density estimation

with some brief references to programming languages and techniques).

While browsing on Internet, I noticed that the Ph.D. courses in this area mostly refer to the term Computational Statistics. The content of the courses fairly corresponds to the mentioned textbooks with some additions regarding most recent techniques like resampling methods and statistical visualization. When the adopted term is Statistical Computing, the relative courses are mainly oriented to meta-languages and statistical software.

I believe the contents of textbooks and Ph.D. courses offer a first idea of what can be meant with Computational statistics, as the result of the interaction between Statistics and Numerical Analysis that favored the implementation of statistical methods as well as the development of new ones, that is, according with B. Efron, to think he unthinkable.

This view of mine is close to what Andrew Westlake (1994) proposed during a debate on the IASC list-server: „Computational Statistics relates to the advance of statistical theory and methods through the use of computational methods. This includes both the use of computation to explore the impact of theories or methods, and development of algorithms to make these ideas available to users.“

Murray Cameron, in the same occasion, offers a Statistical computing "broad view" including "most modern applied statistics as well as some aspects of theoretical statistics (e.g. Bootstrap, MCMC)" as well as a "narrow view" regarding the "Computational aspects of statistical methods such as languages, environments, libraries, packages, to support this work."

This definition, mostly referred to the activities of users of Statistics, neither seems general enough nor takes into account the evolution of the Computer Science tools which offer new and more efficient ways for an effective Statistical computing. Although Computational statistics and Statistical Computing so far argued show some overlappings (see Figure 1) in their possible definitions, as the latter develops from the former, it deserves a more specific definition.

Actually, Statistical computing represents the natural evolution of the traditional computations in the

activity of statisticians. This activity now takes advantage from the new computer and software environments thus increasing the use of Statistics in other fields. The new Computer Science technologies extend Statistical Computing to other basic tasks of Statistics like data collection and dissemination (EDI, Networks), their organization and management (SDBMS), interpretation of results and reporting (AI), disclosing new perspectives for Data Analysis though computer graphics, neural networks and so on.

I therefore prefer to define "Statistical Computing as the application of Computer Science to Statistics", and, on the other hand, "Computational statistics as aiming at the design of algorithm for implementing statistical methods on computers, including the ones unthinkable before the computer age (e.g. Bootstrap, Simulation), as well as to cope with analytically intractable problems".

Figure 1 highlights these definitions in terms of interactions between disciplines showing also a significant area of multidisciplinary research.

It must be noticed that even if the emphasis of the above definitions is on Computer Science or/and Numerical Analysis, the context is always statistical. Obviously, the contours of the previous defi-

nitions can not be sharply drawn but must be necessarily flexible in order to take into account the continuous changes and improvements characterizing Statistics in computer age.

After clarifying the terms of the question, it seems to me that the next COMPSTAT meetings as well as the official IASC journals should refer to the Statistics in Computer Age at large, including, in a proper balance, themes of Computational Statistics and Statistical computing without forgetting the applications to different disciplines according with the mentioned IASC objectives.

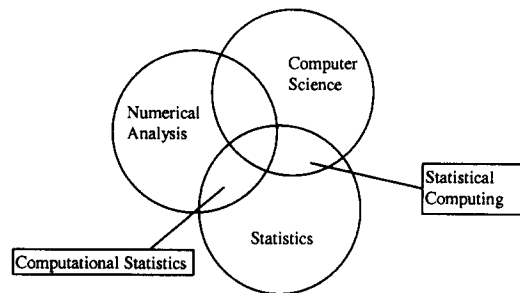


Figure 1: Statistics in Computer Age

* Editorial: Statement presented during COMPSTAT '96