Para discusión en el bloque 1

Tomado del libro "Elements of Computational Statistics" (2002) de James E. Gentle, ed. Springer

In recent years developments in statistics have to a great extent gone hand in hand with developments in computing. Indeed, many of the recent advances in statistics have been dependent on advances in computer science and technology. Many of the currently interesting statistical methods are computationally intensive, either because they require very large numbers of numerical computations or because they depend on visualization of many projections of the data. The class of statistical methods characterized by computational intensity and the supporting theory for such methods constitute a discipline called "computational statistics". (Here, I am following Wegman, 1988, and distinguishing "computational statistics" from "statistical computing", which we take to mean "computational methods, including numerical analysis, for statisticians".)

The computationally-intensive methods of modern statistics rely heavily on the developments in statistical computing and numerical analysis generally. Computational statistics shares two hallmarks with other "computational" sciences, such as computational physics, computational biology, and so on. One is a characteristic of the methodology: it is computationally intensive. The other is the nature of the tools of discovery. Tools of the scientific method have generally been logical deduction (theory) and observation (experimentation). The computer, used to explore large numbers of scenarios, constitutes a new type of tool. Use of the computer to simulate alternatives and to present the research worker with information about these alternatives is a characteristic of the computational sciences. In some ways this usage is akin to experimentation. The observations, however, are generated from an assumed model, and those simulated data are used to evaluate and study the model.

Advances in computing hardware and software have changed the nature of the daily work of statisticians. Data analysts and applied statisticians rely on computers for storage of data, for analysis of the data, and for production of reports describing the analysis. Mathematical statisticians (and even probabilists) use the computer for symbolic manipulations, for evaluation of expressions, for ad hoc simulations, and for production of research reports and papers. Some of the effects on statisticians have been subtle, such as the change from the use of "critical values" of test statistics to the use of "p-values", while others have been more fundamental, such as use of multivariate and/or nonlinear models instead of univariate linear models, which might formerly have been used as approximations because they were computationally tractable. More recently, computational inference using Monte Carlo methods are replacing asymptotic approximations. Another major effect that developments in computing have had on the practice of statistics is that many Bayesian methods that were formerly impractical have entered the mainstream of statistical applications.
The ease of computations has brought new attitudes to the statistician about the nature of statistical research. Experimentation has been put in the toolbox of the mathematical statistician. Ideas can be explored via "quick and dirty" computations. Ideas that appear promising after an initial evaluation can be pursued more rigorously. Larger scale computing systems have also brought new attitudes to the statistician about the nature of discovery. Science has always moved ahead by finding something that was not being sought. Exploratory methods can be applied to very large datasets. Data mining of massive datasets has enabled statisticians to increase the rate of finding things that are not being looked for.

In computational statistics, computation is viewed as an instrument of discovery; the role of the computer is not just to store data, perform computations, and produce graphs and tables, but additionally to suggest to the scientist alternative models and theories. Many alternative graphical displays of a given dataset are usually integral features of computational statistics. Another characteristic of computational statistics is the computational intensity of the methods; even for datasets of medium size, high performance computers are required to perform the computations. Large-scale computations can replace asymptotic approximations in statistical inference.