COMPUTER SCIENCE AND STATISTICS:
The Interface, James E. Gentle (ed.)

MACHINES AND METAPHORS

Joseph Deken

Departments of General Business & Computer Sciences
University of Texas
Austin, Texas

Computer hardware developments are causing radical changes in both the power and the availability of personal computers. In the sophisticated computing environments of these new machines, computationally effective new metaphors are emerging. The static and linearly connected models of traditional mathematics are being augmented with the notion of parallel, communicating processes. Object-oriented programming and an application to data analysis are illustrated.

In a session entitled "Software Trends," it is appropriate to have a variety of horizons. On the near horizon, I'm sure that many participants in this session are directly and immediately involved in using today's microcomputers and today's statistical software, to solve problems that need to be done today (or yesterday). For my part, I would like to scan a slightly farther horizon and discuss trends which may be felt widely within the next five to ten years.

The rapid advance of microprocessor technology we are immersed in represents much more, than simply a new era of convenient conventional computing. What we are seeing is a microcomputer revolution, which promises as a byproduct to initiate fundamental changes in both the theory and practice of statistics. As our computing machines change in both power and accessibility, they are making possible the use of new and more appropriate metaphors for data analysis.

In this paper, I would first like to establish a perspective: What is the personal computing phenomenon, and what are the important aspects of the "new machines?" From that technological perspective, I will then propose a new approach to computationally effective metaphors for data analysis. New approaches will inevitably be developed, to adapt to the radically new microcomputer environment. For the benefit of both our profession and our clients, we statistical professionals should be taking an active part.

The Micro Wave

Even if microprocessor technology meant nothing more than old computers in smaller, cheaper boxes, there would still be important repercussions of this technology for the practice of statistics. Those of us who appreciate practical techniques for real-world data analysis, could hope to broaden our definition of what is "feasible" or "interpretable, even if nothing changed in computing, except that every practicing statistician would be given a personal mainframe computer. We would be able to try more alternatives, do more experiments, and deepen our own insight (and perhaps that of our clients) into both traditional and novel methods of data analysis.

The impact of microcomputing reaches though, not just to us as professionals, but to our clients as well. It would be obviously foolish for statisticians to ignore the availability of their own powerful computers to assist them in conducting their analyses of data. It would be equally foolish to ignore the direct use of computers by their final audience. No longer can we realistically think of the statistician as the anointed emissary who takes the data from the client, goes to the computer, and returns with the truth graven on stone tablets. Thanks to microcomputers, the computer is no longer just the professional's tool, but is becoming a familiar part of the client's everyday environment. Like it or not, both non-statistical professionals and non-professional statisticians of all types are going to have computers. And, wisely or not, they will be using them.

A Changing Audience

Effective data analysis rests on the skillful use of both feasible and interpretable techniques. Rather than simply focusing on the effect of our newfound professional computer in determining what is feasible, we should also assess the impact of the client's newly discovered personal computer on what is interpretable. The importance of both feasible and interpretable procedures is recognized by any successful applied statistician. There is a nice encapsulation in John Tukey's notes on multiple comparisons:

<> Practical power: The product of the mathematical power of a technique and the probability that the technique will be used. (Churchill Eisenhart)

<> Useful power: The product of the mathematical power by the probability that, when a positive result is found ... this result can be understood and clearly interpreted, (partial understandings and interpretations duly weighted.) (John Tukey)

Some of the most interesting work in computing today, in the area of graphics and user interfaces, rests on the simple observation that most people are adept at visual processing and unskilled at speaking mathematics. What I am suggesting here is that in the near future, we ought to add the further, simple observation that most of our clients regularly use personal computers. My contention is that the regular use of a personal computer will become as relevant to the way a person thinks as the fact that the person speaks English say, rather than BNF.

New Machines

It is perhaps obvious, but worth emphasizing nonetheless, that the widespread use of personal computers means much more than it would if these computers were of the programmable pocket calculator variety, or even roughly similar to the familiar eight-bit microcomputers which "broke the ice" in personal computing. The personal computer of the near future will at once be much more powerful and simultaneously much more accessible and easy to use. (Machines such as the Apple Lisa, Xerox 8010, SUN workstation, or Symbolics' Lisp Machine are just a few worthy examples to illustrate the beginnings of this trend.)

Besides containing powerful microprocessors, the modern and future personal computer will be more accessible than ever because of other hardware developments. In particular, the quality of computer displays has made a quantum move forward in the last few years. Rather than seeing a screen which is rigidly divided into a small number of row-and-column positions for alphanumerics, the modern computer user can enjoy a flexible bit mapped display. In the bit-mapped display, the screen may easily have a million independent picture elements. (Of course, any few dozen neighboring elements may often be configured to form a letter or number. In fact, effective text presentations, using a variety of character sizes and styles simultaneously on the screen, are one of the most immediate benefits of a bit-mapped display.)

Far more consequential than its ability to change type faces, the bit mapped display has the capability of being divided into concurrent, interacting windows. These windows add a powerful new dimension to the computer's effectiveness in multi-tasking. Each individual window can display not only text and conventional graphics, but pictorial icons and animated images as well. With such versatile display capability, each window can have a communicative and comprehensible "personality," effectively representing its underlying process to the user.

Communication though, is a two-way street. Hand in hand (or hand-on-screen) with developing graphics, speech synthesis, and other output capabilities, there is a powerful trend emerging toward more flexible input devices. With a modern interface, the human can communicate to the computer by pointing, motioning, touching or even speaking.

When input and output capabilities emerge from the first steps now being taken, into anything like full stride, the personal computer will not be so much a machine as a conversant, powerful array of "companions" able to subtly and powerfully shape people's expectations and thinking styles. An effective computer graphics interface can go far beyond words in making the Euclidean geometric projection metaphor powerful, practical, and directly useful for human data analysts. In the same way, the new style computer, with its multi-faced, responsive and interactive interface, goes far beyond "fossil" imagery in opening up a world of new, actor-like metaphors for understanding data.

For all its beauty and durability, classical mathematics furnishes a restricted source of metaphors for human data analysts. In fact, in the style of mathematics which underlies all of theoretical statistics, there is only one type of "interaction" (function application) and one type of "communication" (function composition). It is a highly serial world, where the symbols are "dead." (The innards of a corpse may be rearranged without apparent effect. The symbols in a mathematical theorem have the same property.) Even the graphics in this mathematical world are "fossils," which can only be manipulated in the same way that a skull could be rotated before pressing into sand. (Hamlet aside, conversation with skulls or fossils is not usually attempted.)

If we could, it would be very effective for us humans to expand our analytic "language" beyond the linear, fossilized realm of classical mathematics, and bring active, parallel, communicating, and conversational metaphors into play. I have tried, in my description of computer technology developments above, to convince you that indeed we can and will develop these new analytic tools. Rather than simply speculating about details of the future though, it would be informative in conclusion to describe just one approach which is being actively pursued, and already produced effective new computing metaphors.

Object-Oriented Data Analysis

There are several different versions of object-oriented programming actively under development, though much of the credit and early work is associated with the Smalltalk group at Xerox Palo Alto Research Center. The details of implementing an object-oriented system can vary considerably, but the underlying concept is constant and most relevant here: Unlike a conventional serial programming language, in which function application is managed by linear control structures, the object-oriented system seeks to establish parallel, communicating agents. These agents are highly modular and independent, and communicate both among themselves and with any human participants. (For example, Smalltalk is organized in terms of actors, methods, and messages. Upon receiving a message, the actor uses one of its methods to take action.)

In an object-oriented environment, humans can use the computer effectively by being good "managers" and good "directors," as well as logicians or mathematicians. The computer is not furnishing functions it seems, as much as a competent and co-operative staff of "assistants." The key is to organize the given agents' capabilities efficiently, to communicate with them effectively.

As a specific example, we can consider the classic case of "regression." (Actually, I have used the word loosely. What is described here could be a framework for doing either familiar types of analyses or entirely new investigations.) I will describe three agents, describe their methods, and indicate their paths of communication. The "staff" structure created here would effectively and interactively assist a human data analyst to understand a quantitative situation.

The Data-Item Agent: There are many of these agents, all with the same methods and intercommunication pattern. Each data-item agent is associated with an observation. The capture methods handle such tasks as data acquisition, verification, missing-value reconstruction, and so on. The compromise methods handle "fitted" approximations to the observation, "voting" with some weight for that observation's "position." The group methods enable an observation to find and communicate with "neighboring" observations, should that be relevant. The communications of a data-item agent are directed to its neighbor data-items, to any active organizers (agents described below), and to any reporters (see below also).

The Organizer Agent: Any attempt to make overall "sense" of the observations is handled by an organizer agent. In the give and take between the organizer and the individual observations, the organizer's methods might be to poll the data-items as to their preferred positions, suggest compromise positions, inform data-items of non-local conditions, and ultimately arbitrate perhaps a final organization.

The Reporter Agent: The reporter might function as the principal agent to manage the human/data interaction. The state of the analysis and the current interactions between the other agents could be presented or recast in a variety of sensible and sensory formats. The reporter might also need methods to interpret questions from the human analyst, or to generate relevant questions for consideration.

Here is a tabular summary of the "staff." It is only a suggested cast of characters. In the dramatic era ahead, we can look forward to many casts, many plays, and much insight:

agent	methods	message
Data Item	capture compromise group	neighbor organizer reporter
Organizer	poll-suggest inform arbitrate	data items reporter
Reporter	present recast interpret -questions	organizer data items