Statistics and Our Expanding and Shrinking World

L. Billard
Department of Statistics, University of Georgia
Athens, GA 30602 U.S.A.
lynne@stat.uga.edu

As I sat down to begin this article, the darkening light outside suggested a storm was imminent. Therefore, I turned in to the webpage whose local doppler radar confirmed that heavy rain was moving my way. The same doppler screen serendipitously featured a banner proclaiming the availability of "Forecasts from Afghanistan to Zimbabwe". In an instant, I learnt that Kabul was windy and Harare was experiencing storms. This example might on the surface seem trite, but in fact it is fraught with symbolism at several levels of the argument. No matter where we sit, be that in America, in Asia, in Africa or wherever, with a keystroke or two we can bring the whole world into our immediate environs, reinforcing the notion that the new technologies are shrinking our expanding world rapidly. This assumes we are connected to this modern electronic era. If not so connected, then the modern world (and all this phrase implies today) is exploding so fast that it is the individual that is shrinking, becoming lost in the vastness of this expansion. Despite the contemporary changing modes and speeds of communication, and though we are aware that not all of us are presently connected, it is important for all of us to position ourselves so that we do move into this future and not be left behind. Fortunately, it is my firm belief that we will all be connected (probably sooner than we think, or care to contemplate for some!), though also right now we sit in the unsettled interim as the new world emerges.

Whilst these remarks imply a physical world defined by geographical boundaries, it is in reality the world of knowledge that is the focus of our thoughts herein. Knowledge itself knows no boundaries, nor are its components limited by distances but rather they travel to whatever corner bids its calling, impelled only by the power of intellectual curiosities.

As we adjust our intellectual telescopes, we quickly realize we have first to wrestle today with a technology, symbolized by a superhighway, that was not possible as recently as 15-20 years ago even though the key ingredients had been discovered as far back as the 1940s. (See National Science Foundation, 1994, for a fascinating article on the origins of the information superhighway - a powerful endorsement for the pursuit of basic knowledge incidentally). Nor could we dream perhaps even 10 years ago just how much the figurative office walls would be torn down, so that today we can interact collaboratively with colleagues anywhere in the globe much more easily than in bygone days. As a consequence, the basic infrastructures of our operational worlds have to be reconfigured in vastly new ways.

The superhighway connects the world - the world of individual scientists, libraries, schools, institutions, governments, etc. - into one giant classroom where we can all listen to and hear the same information delivered by the one (same) instructor. As such, we have, in theory, an avalanche of material (knowledge) available to us as never before (without having to move outside, without having to travel to those knowledge sources.) Equally, just as the student who skips or sleeps in class, if we are not tuned in or are not paying attention, we miss the delivered message(s). Or, equally problematical, we may be engaged but are confronted with so much material we are paralyzed at sorting out which trail we should follow - in much the same way that when asking the time we are told how to build a clock, i.e., there is much more information than necessary.

This assumes all data are accurate, which in fact we know is not the case. By "data" here we mean information generally. As computer power and capacity explodes, issues such as data integrity become
important. This also includes the reliance we can place on scientific "papers", as researchers (and nonresearchers, in fact, anyone) place their unreferred technical reports or articles on webpages. Policy issues regarding the use and sharing of such received information, attendant liabilities, etc., still have to be resolved.

Issues of management of web-data will likewise make their voices heard. As we are flooded by the apparent riches available, we desperately wonder: How can we be sure we have access to all the crucial information relevant to a given situation? And how can we absorb it? When publication outlets were few in number, we could confidently feel we had successfully scoured the relevant literature, and then move forward with our task, and in a timely manner. Today, few of us can be so confident unless we spend more time than is conceivably reasonable. One consequence of this phenomenon is that as the knowledge base and its availability expands, workers shrink their worlds of functional operation into even more narrowly defined specialties. Ironically, this coping action by its very nature starts to rebuild (new) disciplinary walls, thus increasing the (professional knowledge) distances between us.

This doesn't solve the question of finding the material itself. Libraries in the 1800s responded to its version of this issue by developing catalogue systems and by arranging books on shelves in accordance with these systems. Electronic equivalents helping us to navigate the sea of information available do not yet exist, as far as I can tell. Current internet search engines do not address this general issue.

**Electronic library and e-publishing**

As we consider the electronic library and its companion electronic publishing, let us restrict our thoughts to traditional library concerns, viz., books and journals. In contrast to the publication of magazines, leaflets, promotional material and the like (for which it has long been assumed that these will quickly become electronic print material only), librarians at the forefront of the electronic revolution of their trade even quite recently were reasonably convinced that books (especially scholarly tomes) and journals would continue to coexist as both paper print and e-print for longer than most visionaries suggested a few (five?) years earlier. The logic here was that page(s) of such volumes by their very nature would be read and reread, digested and pondered, lovingly savored by the scholar and so forth, all necessitating the traditional paper copy. However, it should be noted that in drawing on a variety of sources, *Nature's* recent briefing on electronic journals (see Butler, 1999) paints a somewhat different picture suggesting that "while high-circulation journals will exist in print and electronic versions for the foreseeable future, the low-circulation, high-cost journals that make up the bulk of the scholarly publishing market will in future exist only electronically". This opinion derives from the wholesale movement of traditional publishers to the internet, and from their bypassing libraries by providing all their journals directly to the user.

Many questions abound. Publishers (which includes professional societies, such as our own, who publish journals) wrestle with issues of production costs (on the whole these are approximately equivalent for electronic and for hard copy versions of the same publication), costs to the electronic user, how to make what material available (e.g., should there be an electronic equivalent of a pay-as-you-use zerox machine, or what?) and when. Copyright, archiving, legal and ethical issues must be satisfactorily resolved. These are quite difficult issues, and can be dauntingly complicated; witness the prolonged debates over the WIPO intellectual property rights legislation. Likewise, archiving (outside of the paper version) has no immediate solution, especially when technologies keep changing so rapidly. In contrast, scanners allow past volumes to be immediately accessible (e.g., entities such as JSTOR are now embarking on e-archiving entire sets of journals). For the scientist, perhaps the most crucial issue is the quality control derived from peer review. The *Nature* briefing sees the main role of journals as providing the imprimatur of quality and editorial value with their traditional role as a distribution outlet receding. While it seems journals are fully destined to have electronic twins, and
likely followed by monographs sharing this same rite of passage, it is not yet clear if people will read entire books online.

Obviously without the technological advances, none of this would be possible. It is also true that economics, as in declining library budgets, is hastening this movement. Where economics in the past has divided our worlds, be these divisions between relatively well financed libraries in major institutions or poorly financed small college libraries, or divisions on geographical, political or cultural lines, the electronic library will help erase these boundaries, to anyone with access to the internet. This is an enormous advantage. I well remember visiting the library of what had been a major institution in a developing nation, but which because of internal problems could no longer support its research and educational/infrastructure so much so there appeared to be no volumes of any description that had been published after the early 1970s. Yet, there was email and internet access, which access if developed appropriately will help close that gap considerably. In fact, it is often the case that electronic access (albeit perhaps still only intermittent such as once-a-day but improving, rather than instantly on demand) is more reliable than surface mail and is certainly faster. Expansion of these capabilities will most surely help shrink the knowledge gap in both directions.

Training

Training involves retraining ourselves and training our students. In either case, there is the need for training which allows growth, adaptation to the every-changing world, innovation, and so on, over a 30-40 year career span. We all have to learn new technological skills. Those of us who completed our formal training before these certainties dawned, have the added task sometimes of overcoming mental hurdles. This requires changing our way of thinking to "the possibilities of tomorrow instead of the boxes of today"; in the doing, everyone becomes empowered and energized as the new world expands our horizons. What we can achieve in collaboration with others will be enhanced and accelerated as we embrace and adapt to these evolving technologies.

Having (re)trained ourselves, we can then more effectively train our students, both in the "what" and the "how". Changes from the traditional classroom approach press upon us. Cyberspace instruction and its myriad facets are today in an analogous position to that of publishing a few years ago, typified by Crews’ (1996) comments that enlightened confusion abounded with policies that were elusive, ambiguous, hopelessly vague and continuously perplexing, with little consensus forming on how to proceed. As we wrestle with installing web instruction, we must not forget that, in addition to coming to terms with the technology, we have to recognize that it will require budgetary resources yearly, faculty will need to be trained, and it will not necessarily save money but rather will allow us to do things better and in new ways with immeasurable benefits. We will be able to reach new audiences, without being fettered by time or place. It is doubtful however that the intangibles transmitted by face-to-face exchanges can ever be successfully replaced. Rather, cyber-instruction will more likely serve as a complementary tool to the total learning process. (See, e.g., Hahn and Jackson, 1995 and Rudenstine, 1997).

Research

The foregoing remarks apply wherever we are stationed in our professional lives. For us as statistical scientists, computers will play an additional and fundamentally substantial role. In one direction, for problems requiring known methodologies, computers ease the computational burden of conducting statistical analyses, especially for moderate to large data sets. A concomitant issue is the use of statistical packages, welcomed additions to our toolkits. Unfortunately, these packages also bring new concerns, the most important of which is their blind use, largely by researchers in nonstatistical disciplines, whereby an inappropriate routine/analysis is applied to the data at hand.
Another concern revolves around the lack of proper direction in the accompanying manuals. For example, as those who use such packages frequently are painfully aware, oftentimes following manual instructions will give incorrect answers.

In a completely different direction, theoretical developments were previously restricted to situations that could be reasonably handled in "noncomputational ways" (though not discounting the aforementioned computer use). New methodologies that could not have emerged without the power of the modern computer include the analysis of large and very large data sets using for example symbolic data analysis and data mining methods, statistical graphics, resampling methods, imaging, encoding-encryption methods, and so on. This is another whole topic in and of itself. What is clear is that there is an urgency to our own discipline to be engaged fully and fruitfully in these areas. Some, such as data mining, are already being syphoned away to other disciplines. It is imperative that we respond responsibly to these effects.

Conclusion

It is clear the electronic revolution is producing more tools for us to access more information to enhance the pursuit of scholarship. There will be an explosion of information at our fingertips with electronic links to reference materials, databases, and there will be graphical innovations such as animated displays and working examples of these materials. With each new development our expectations increase. Just as knowledge expands, so too do our audiences grow. Every facet of our professional lives will be intricately entwined with computers. The pace of change will be rapid and profound and just as we conquer one mountain another awaits us; but master it/them we must. Our task is to be alert to the possibilities and then to develop these potentials into useful adjuncts (or perhaps replacements) to the central tenets of our science.

References


RÉSUMÉ

Le décon d’excercer notre profession. Nous regardons quelque avantage de cette situation, particulièrement en ce qui concerne la publication, l’enseignement, et la recherche scientifique.