Statement of Professional Goals and Objectives

Ranjan Maitra

Introduction

The past decade has seen the growing importance of statistical thought in providing practical solutions to real-world problems. The engineer, the scientist, the medical researcher, the policy-maker and the business executive all want the statistician as a partner in the effort to address vexing issues of practical significance. At play are several levels of statistical expertise – of the experimenter, the analyst or the practitioner. The need for the researcher is then crucial, but no less is that of the educator who nurtures and trains statistically sound scientific lines of inquiry, or of the professional providing valuable service in advancing the goals of the discipline to make research and education possible. In this document, I illustrate my specific goals in each of these areas, my attempts towards realizing them as a member of the faculty, the university and the statistical community, and my vision for the future.

Research Objectives

Technological and scientific advances have resulted in automated and efficient data collection methods. They have also thrown up interesting research questions in the information, life and engineering sciences. Classical statistical methods have proved to be deficient for ready application in many cases, especially in the context of massive databases. My primary research interests continue to be in the development of statistically sound practical methodology to address the challenges in such scenarios, with special emphasis on spatial statistics and image analysis, bioinformatics, data mining and the environmental sciences.

My initial exposure to problems of this magnitude was in the context of image analysis, as a graduate student at the University of Washington. Of interest there was the development of inference tools to guide clinical diagnosis, using *functional images* which are comprehensive displays of tissue biologic parameters, *e.g.* metabolic rate. An assessment of the variability inherent in the process of data collection is important and my research provided a first step in this direction. A related manuscript was identified in 1998 by the journal *Medical Image Analysis* as one of eight forward-looking papers in a special invited issue on information processing in medical imaging. In recent months, I have shown that one can rewrite discrete multi-dimensional Fourier Transforms in real form, yielding a practical way for generalized cross-validation (GCV) estimation of the optimal bandwidth of standard Positron Emission Tomography (PET) image reconstruction. I have also been collaborating for the past several years with scientists from the University of Maryland Medical Center and Johns Hopkins University in functional Magnetic Resonance Imaging (fMRI). The goal of fMRI is to understand cognition, *i.e.* to locate and understand areas of activation in the human brain in response to different tasks or stimuli, and to eventually use this information to detect anomalies in pathological cases. The statistical challenges posed here are at once both classical and modern: for instance, identification of voxels showing significant activation is the age-old multiple testing problem severely compounded by the fact that only at most 2-3% of the almost 100,000 voxels imaging the brain volume are expected to be activated. To this end, I am directing the Ph. D. dissertation of William Baumann in developing and exploring a new adaptive approach. This particular problem also rears its head in areas such as the analysis of microarray data in *bioinformatics* or in network intrusion detection in *telecommunications*. In the fMRI context, issues arise not only from the size of the three-dimensional images collected over time but also from the mode of their acquisition – there is substantial spatio-temporal movement as well as noise due to motion. To complicate issues, some of the motion is paradigm-induced (for example, the application of a moving stimulus in a person's field of vision tends to induce reflexive action) while some of it may be due to unrelated causes. New statistical methodology is therefore needed to quantify the sources of variation inherent in such studies *en route* to the more important goal of understanding the cognitive processes. The next step is to establish whether the variability in activation due to physiological condition is more than that due to inter-subject differences. In the long term, our plan is to use this to understand the degenerative effects of neurological and other disorders with a view to facilitating improved patient care and therapy. The focus of our work has been very quantitative, with results published in several flagship statistical and medical imaging journals. Further, I have continuing funding from the National Institutes of Health to support my student's research as well as part of mine for five years from 2004-09.

The issue of finding clusters in data is a difficult problem that is intractably compounded when the dataset is massive. This is an issue of interest not just to statisticians but also to researchers from several fields, including very lately, those from the computer and information sciences. I have provided a multi-staged iterative approach — my initial research on this subject was published in the *Special Tukey Memorial Issue* of *Technometrics* in August 2001. The original research was in the context of telecommunications software metrics data partially funded by a set of two external industrial research grants (1999, 2000-01), but this was later used to understand self-reported United States Environmental Protection Agency (USEPA) data on toxic releases of different industrial facilities. Several aspects and offshoots of the clustering problem demand attention: I recently submitted manuscripts on efficient simulation methods for generating multivariate data from extreme regions and on deterministic initialization methods for iterative clustering algorithms. Additionally, I am directing the dissertation research of Ivan Ramler in investigating methods to assess robustness in clustering methodologies and also in segmenting indirectly observed data – this is important in a wide range of applications, notably dynamic PET imaging where the goal is to obtain a comprehensive motion picture display of human organ function and biology. Another area of interest is the clustering of multiple data-types. This is important in a wide range of applications, its need being increasingly felt in the context of microarray analysis in bioinformatics. In the long term, I also intend investigating clustering strategies for data streams and for datasets that are incrementally available by coordinate. Such methodology is very necessary in the context of financial and profile data, of which there is now a surfeit because of the increasing number of internet-based services.

A third long-term research area for me is the development of new approaches to stochastic computation. I have a continuing interest in the development of parallel simulation and estimation approaches to Markov Processes and Markov Random Fields (MRF's). For instance, the continuum Markov cluster process is not a pair-wise interaction process but is still plagued by the phenomenon of low mobility in stochastic simulation. My first dissertation advisee, Dr. Rafaela Guidi and I developed and investigated *auxiliary variables* schemes for these processes. The cluster process is a good choice as a prior object model in the analysis of two-dimensional proteomic gels and efficient simulation approaches are needed to obtain estimates in this context. In general, a disquieting feature of simulation methods using auxiliary variables is the large numbers of additional variables often needed — this can be quite a daunting requirement in severely multi-dimensional applications such as fMRI. I am interested in developing the use of multi-grid simulation methods, through the introduction of parallel MRF's at coarser scales — which because of their fewer coordinates have greater mobility. A challenge here is the amalgamation of the different scales while keeping the marginal distributions at each scale straight-forward. Another area of interest to me is the development of methodology for estimation in MRF's with spatially varying interactions. I intend looking into an adjustable grid which can be modified such that interactions in the new grid are spatially invariant. Wavelet techniques are proposed to speed up computations in the estimation process.

Many of the research issues presented here arise because of the rapid advances in technology. My current and planned solutions are built on exploiting computational resources such as modern workstations, parallel virtual machines and supercomputers. These need to be harnessed in a smart and efficient manner. Most importantly, I believe that the research challenges here will only become more acute with technological advances and more efficient and automated data collection methods. Thus, the significance of my research objectives extend well beyond the immediate needs and will potentially be relevant for a long time. Indeed, the plans outlined in this section form substantial parts of the funding award mentioned above as well as my continuing five-year (2003-08) CAREER award titled "Methodology for Statistical Computing in Massive Datasets: Parallel Approaches to Clustering and MCMC Estimation" funded by the National Science Foundation (NSF).

Teaching Objectives

The primary role of an university is to disseminate knowledge and to prepare well-equipped citizens for society. It does this by fostering analytical thinking and by preparing students for the many facets of an increasingly technologically-driven lifestyle. Statistical education is no different, whether in the undergraduate classroom, the graduate level or in a mentoring context. This principle has always been the cornerstone of my teaching philosophy at both the University of Maryland Baltimore County (UMBC) and at Iowa State University, even though the emphasis has been modulated for each student by his/her exact needs and educational background. I consider teaching to be a very fulfilling experience and have myself grown as an educator over the years.

I believe very strongly that undergraduate education is the foundation for a student's future in any discipline. At the same time, graduate classes develop in our students the expertise to practice statistics and to extend the frontiers of research. During my appointments, I have taught introductory undergraduate classes in statistics for social, engineering, pure and the life science majors. While doing so, I have tried to relate my classes to my students' major disciplines with examples from day-to-day life and relevant applications. I have also taught upper-level undergraduate classes in applied statistics and in statistical software in addition to designing an undergraduate special topics class in data mining and in statistical computing. Most recently, I have developed an undergraduate research experience class, to be offered to incoming freshmen interested in statistical research at Iowa State in Spring 2006. At the graduate level, I have taught core courses in applied statistics and designed and taught more advanced courses in multivariate statistics, nonlinear models, statistical computing and spatial statistics and image analysis. In both my undergraduate and graduate classes, I include several real-world data analysis project assignments, with professional-grade reporting requirements that emphasize clear communication, consistent with our society's need for promoting writing across curricula. In my advanced-level graduate classes, the assigned projects are intended to provide students with a full research experience: indeed, some of of them have eventually submitted their project reports to professional journals or meetings. Student evaluation scores on my teaching have been very high – in particular they have mostly been above 4 (on a scale of 0 through 5) over the past five years on the key questions on preparedness and effectiveness, and indeed above 4.5 in many of these cases. In addition, the Department of Statistics at Iowa State has a mechanism whereby a senior professor visits a junior faculty member's classroom for two consecutive classes every 3-4 years to provide detailed feedback to the chair. Professor Max Morris visited my class in multivariate statistics in this regard in Spring 2004 and provided a very positive evaluation.

My teaching responsibilities have provided me with the experience to contribute effectively to both the development of curricula in different areas of statistics (in statistical computing, data mining, spatial statistics and image analysis). I was very closely involved with the successful effort to develop Maryland's first statistics major and minor programs at UMBC. More recently, I was charged with leading the effort to recast the graduate curriculum in statistical computing at Iowa State. The new sequence has just completed its first cycle and its content has been viewed very enthusiastically, both by faculty and students. I have also been very active in the mentoring of graduate students. I was the dissertation supervisor for Dr. Rafaela Guidi and am currently supervising the dissertations of Mr. William Baumann and Mr. Ivan Ramler. Besides supervising Masters' "creative component" research of three students, I have mentored undergraduate and graduate students as part of the NSF's VIGRE-funded project at Iowa State – in this context, I should mention that I developed and continue to lead the statistical computing component of the VIGRE program. I am a senior personnel in the department's successful Research and Training Graduates (RTG) funding request to the NSF for supporting graduate students. These have been very exciting and learning experiences for me, as I have been able to combine teaching and mentoring with the thrill of extending research frontiers. I believe that these experiences will be very useful launching pads for the challenges ahead. Additionally, my research objectives gel well with the requirements of new graduates in an increasingly technological world. For instance, an increasingly substantial part of modern statistical research involves scientific computation and I believe that we need to continuously fine-tune and update both undergraduate and graduate instruction to keep abreast of such trends. This can be augmented through the development of suitably-tailored classes as well as through hands-on research experience projects. I consider myself well-equipped and experienced to develop and re-design such classes which will use both traditional and more modern methods of instruction.

Service

An integral part of my professional responsibilities has been service to the discipline, the university and the community. I have been a member of the Departmental Computer and Undergraduate Committees and the University Faculty Instructional Committee. I have also played an active role in the recruitment of undergraduate and graduate students. As indicated earlier, I have been involved in efforts to develop programs and curricula. I believe that my experiences here will come in useful as new programs and courses are needed to be developed and existing ones modified. I coordinated the department's successful effort to get NSF funding under the SCREMS program. At the professional level, I have reviewed manuscripts submitted by my peers for publication in several journals. I was appointed to the Executive Committee of the American Statistical Association's (ASA) Section on Statistical Computing in 1997 and stepped down this year after eight years of service. I have also been Continuing Education Liaison for the same section (1997-2001) and the section's Editor of the Statistical Computing and Graphics newsletter (2002-05). I have been an elected President (2000-01, 2001-02) and Vice-President (1998-99, 1999-2000) of the Maryland Chapter of the ASA. At the community level, I was involved with the ASA's Adopt-a-School program which enables professional statisticians to help schools in their efforts to bring quantitative literacy into their classrooms. Thus I believe that my active service to the department, university, community and profession has been an enriching experience and has prepared me for further challenges ahead.

In conclusion, I believe that my research, teaching and service goals are very important and relevant to the needs of modern society. Indeed, they sit well with the cross-disciplinary flavor of modern statistical research and are in tune with the professional interests of researchers across many disciplines. These are very exciting times for statisticians and my past professional experiences and achievements provide me with the confidence that I can contribute effectively in the onward evolution of society, while also developing further as a valued scholar, teacher and member of the community.